A SPATIAL DECISION SUPPORT SYSTEM FOR AUTODISTRICTING COLLECTION UNITS FOR THE TAKING OF THE CANADIAN CENSUS

SIDNEY WAYNE WITIUK
B.Sc., M.A., M.Sc.

Ph.D.
UNIVERSITY OF EDINBURGH
1990

DECLARATION

This is to declare that this thesis is my own unaided work and has not been previously submitted for any degree in any other University.

18 May 1990

date

DEDICATION

To Kathleen, Karole and Kelsey.

ACKNOWLEDGEMENTS

Although it is not possible to name each and every individual who contributed in some important way or another to the completion of this dissertation, I wish to begin by thanking, in a general way, the clerical, secretarial, technical and professional staffs of the Department of Geography at the University of Edinburgh and of the Geocartographics Division at Statistics Canada for their eager and timely support.

During the planning stages, the work was shaped and the directions strongly influenced by my initial supervisors, T.C. Waugh and Dr. M.P. Atkinson. It also benefitted considerably from the input of Professors Tobler, Steiner, Poiker, and Aangeenbrug. Professors Wrigley, Openshaw and Coppock identified particularly useful literature and personal contacts.

The early development stages were particularly influenced by inputs from districting specialists and their supervisors in the Geography Division and by the end users in the Survey Operations Division at Statistics Canada. Useful information about piece rate calculations and time and motion studies for earlier Censuses was made available by staff in the Social Survey Methods Division. I am also indebted to R. Marx, F.R. Broome and B. Rosen of the Geography Division at the U.S. Bureau of the Census for their ongoing interest and encouragement and for their willingness to openly share the relative strengths and weaknesses of their autodistricting efforts.

The results of this research would not have been as conclusive, nor as impressive, if it were not for the support of Census Management to fully explore the potential of autodistricting for the 1991 Census of Canada. I am particularly grateful to Dr. E.T. Pryor, D.R. Bradley, and E. Outrata for their personal involvement in reviewing the approaches and methodologies and, subsequently, for their strong commitment to a full implementation and testing of pilot, prototype and production autodistricting capacities.

Documenting the results of the research has proven to be a major hurdle that was made much more manageable because of the excellent direction and advice of my supervisors Dr. R. Hodgart and Dr. M. Goodchild. Their unselfish and timely review of earlier drafts have dramatically improved the clarity and structure of the dissertation itself. I would particularly like to thank Dr. L. Bondi of the Department of Geography and Dr. G. Deecker, L. Li, P. Tallon, R. Molnar, S. Brockwell and D. Nyman of my staff at Statistics Canada for their useful comments and suggestions on earlier drafts of the dissertation.

I also wish to thank the executive management of Statistics Canada for their granting of education leave and general support of this undertaking.

Finally, I wish to express my deepest appreciation to my wife, Karole, (also a geographer and a fine editor), for her constant support and valuable input. I recognize and greatly appreciate the many sacrifices that she and my daughter, Kelsey, have made in order that this goal could be realized.

ABSTRACT

This dissertation documents the districting requirements for collection units for taking the Canadian Census and provides a spatial decision support system for their automatic creation. In the context of the literature on autodistricting, this problem falls under the general category of creating districts for monitoring, surveillance and inventory applications since the Census is essentially an spatial inventory exercise. The basic requirement is to create an area-based categorical coverage such that the workload is equitably distributed amongst Census Representatives within the limits of a large number of constraints and conditions.

A new omnibus automated districting process that combines a 3-stage cascading selection procedure for identifying sub-blockface, blockface and block level collection units with a 4-stage heuristic solution procedure for grouping blocks (termed 'assigns', 'annexes', 're-assigns' and 'adjusts') is contributed by this research to provide a systematic response to varying districting situations.

The resulting spatial decision support system for autodistricting has been tested on test data sets and on one of the larger urban population centres of Canada. The set of test pattern sites mimicking typical settlement patterns was generated to ensure that the various alternative assignment or block grouping methods (i.e., unidirectional and bidirectional tessellations based on circular and rectangular grids and regular, random and 'extrema-based' seeds) performed as designed and specified. The Census Subdivision of Laval (in the Census Metropolitan Area of Montreal) was selected as the test site for comparing the performance of the autodistricting capacity to the actual, manually created, results from the 1986 Census.

To permit the comparison of results from classical manual and automated processes, a set of satisficing evaluation functions that vary in accordance with data availability was implemented in the context of a competing set of districting objectives. The most sophisticated of these evaluation functions incorporates a composite index that combines the distribution and a measure of the 'density' of the dwellings with the length of the route that must be followed to complete the collection activity (including travel time to the start of the route and between route parts).

To assess the continued acceptability of the districting from the previous Census, and/or to select between alternative results generated by computer-assisted approaches, a set of objective functions is provided that vary depending upon the available amount of geographic, cartographic or statistical data.

Finally, results from some initial testing of method selection mechanisms are reported. Operationally, these have been abandoned in favour of pragmatically employing the three most successful assignment methods in sequence (as required).

In summary, this research provides a clear statement and formalization of the districting process as it applies to the creation of collection units for collecting Census data in a variety of circumstances and of data availability in Canada. An adaptive and flexible model and toolkit of techniques have been developed and implemented as a highly automated spatial decision support system that incorporates the underlying capabilities of a standard geographic information system, ARC/Info. They have also been demonstrated to be effective for the case study areas considered and are being investigated for use in several other applications.

Indeed, subsequent to the completion of this research, the demonstration/prototype system was converted to a production system and used by Statistics Canada to efficiently generate approximately 10,000 collection units, or about 25% of the national total, for the 1991 Census of Canada. Additionally, the system is being assessed for probable application to the creation of "clusters" for the Labour Force Survey, "weighting areas" for attributing characteristics from the 1 in 5 sample of the 1991 Census to the entire population, and "tiles" for storing groups of map sheets for the entire country.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ILLUSTRATIONS

PREFACE

This dissertation documents the results of a research and development undertaking begun in 1981 and sponsored by the author's employer, Statistics Canada. The topic was one of six long term corporate goals identified by the author as part of a successful request for education leave. While there was not a specific requirement from the University of Edinburgh nor from the sponsor that the results of the research be implemented as an operational spatial decision support system, the author was committed from an early stage to demonstrating both the feasibility and the viability of the solution procedure. Because of the sponsor's decision to investigate and (eventually) to adopt the results of this research as a means of improving the cost-effectiveness and the quality of the 1991 Census districting process, and because of the author's functional responsibility as the Director of the staff from the Geocartographics Division that converted the author's specifications and initial capacities to pilot, prototype, and production systems, it is important to clearly specify the role and contribution of the participants through the various stages of research and development.

The research and development effort can be viewed as having four distinct stages:

1. Identification, research and analysis of the problem,
2. Specification, development and enhancement of a solution procedure,
3. Re-implementation, testing and modification of the solution procedure,
4. Converting the solution procedure into a production system capacity.

The last of these four stages did not involve any important changes to the basic solution procedure and, consequently, did not require much technical input from the author. For the earlier stages it was considered to be important that the author's role and responsibility for intellectual leadership be clearly evident to all observers and participants. Except as noted below (or in the body of the dissertation), all of the intellectual and scientific contributions are the sole responsibility of the author.

The following steps were taken in order to achieve the research objective:

1. existing manual districting procedures were studied by the author through a review of available documentation, interviews with districting specialists and "hands-on" training.

2. the literature on districting was reviewed by the author to identify common districting objectives, approaches, techniques and evaluation criteria;

3. the staff responsible for early attempts at computer-aided districting at the U.S. Bureau of the Census were interviewed by the author;

4. a suite of districting tools was developed and implemented as a set of FORTRAN subroutines by the author since none of the existing tools was able to respond to the varying degrees of data availability and the complex interaction of the constraints;

5. the solution procedure controlling the use of the individual districting tools was implemented by the author as a sequence of control language commands (DCL) on VAX computers at the University of Edinburgh and at Statistics Canada;

6. the solution procedure and the districting tools were tested using artificial data composed by the author and using actual data supplied and converted to a suitable format by staff at Statistics Canada;

7. the solution procedure was re-implemented and imbedded within an existing geographical information system, ARC/Info, by Statistics Canada staff under the direct supervision of the author;

8. a set of data dependent objective functions that measure the relative quality of individual collection unit districts were developed and calibrated by the author (after discussions with districting experts at Statistics Canada) to automate the assessment of the districting results and to halt the solution procedure when a satisfactory result is attained;

9. a set of dispersion indices was developed and later extended by the author to guide the selection of alternative districting tools. Testing of the method selection procedure was delegated to Statistics Canada staff once it was determined that using a subset of the method in a pre-defined sequence had a high likelihood of identifying at least one of the best three methods for a given area. As noted later in the body of this dissertation, one of the dispersion indices was developed collaboratively with a member of the testing team;

10. under the direction of the author, empirical validation testing was conducted on each districting method. Shortcomings identified during the validation testing stage and later during the feasibility testing stage were overcome through the incorporation of 'bonding' and 'annexing' techniques specified by the author and implemented by Statistics Canada staff;

11. a set of evaluation functions were developed by the author and implemented by Statistics Canada staff to report on the acceptability of alternative approaches to districting a given set of areas;

12. a set of quality measures were developed by the author to compare the quality of sets of districts produced by the system with the actual, manually generated results from the 1986 Census.

13. at this point, the solution procedure, referred to as a **pilot system** at this juncture, was turned over to an evaluation team under the direction of Ms. P. Tallon - but involving the author - to determine whether or not it could be enhanced to **prototype** level (primarily involving additional input and output capabilities) in time to be used for the 1991 Census;

14. the favourable assessment of the resulting **prototype** system led to the decision to enhance the system to the level of a **pre-production** system in order to assess the economic viability of replacing the proposed manual system with an autodistricting capacity. The author's involvement became less and less direct as the focus of attention shifted to the need for careful planning and integration of the two (manual and automated) production processes. The author worked closely with Task Manager, Ms. P. Tallon, in the area of cost and quality assessment -- with the author focusing primarily on quality and Ms. Tallon focusing on cost considerations.

15. when the decision was taken by Census Management to utilize the resulting **production system** for the 1991 Census, the author's involvement shifted entirely to the managerial level and any minor changes to the solution procedure to facilitate production objectives were considered to be outside the scope of the dissertation undertaking.

The re-implementation team comprised of R. Molnar, D. Nyman, S. Brockwell and G. Lalonde who not only re-implemented the basic capacities (originally programmed in FORTRAN by the author) but did so without submitting to the normal human tendency to embellish and refine the author's methods so that the separation of responsibility would remain unambiguous.

The evaluation team headed up by Ms. P. Tallon did likewise until the system was being converted to a production system (any minor exceptions to this are noted in the body of the dissertation).

While outside the scope of the dissertation undertaking, it is useful to note that the production team, by their dedicated efforts, conclusively demonstrated the flexibility and economy of the design and development philosophy used to construct the resulting spatial decision support system.

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

Districting is the process of classifying or partitioning a specified territory of geographical space into a set of subspaces of identifiable extent, according to a prescribed set of conditions, and is a fundamental activity in geography.

Large scale field operations such as censuses, surveys, mail or flyer delivery, garbage collection, meter reading, police surveillance, and the creation of electoral registers require geographic space to be organized into small districts that cover the entire space and are of a manageable size. Often, these districts must be modified to reflect changing conditions.

Decennial censuses of population have been held in Canada since 1871 and quinquennial censuses of population since 1956. To collect, process and disseminate the results of these censuses, small units for areal collection are defined for the entire country using manual districting techniques.

The set of collection units used in a census has a direct bearing on the quality, cost, and timeliness of the collection, processing and dissemination of the Census. Thus, improvements in the quality and consistency of collection unit districtings can have a direct impact on the quality (in terms of over and under coverage) and usefulness (in terms of timeliness and comparability) of the census data.

The manual process of creating collection unit districts for the Census of Canada is relatively costly and takes over three years to complete. While current time-consuming and labour intensive techniques have been the accepted norm for providing districtings, little is known about the true quality and effectiveness of those districtings.

The development of new methods and measures for determining and improving the quality of collection unit districtings is particularly important given the growing need for increasing the cost-effectiveness of the Census and the importance of the contribution that the districtings make to the timeliness, cost, and quality of published Census results.

1.2 THE RESEARCH GOALS

The principal goal of this research is to investigate the possibility of developing a computer-based model and methodology that can effectively produce (with little human intervention) districtings of comparable or superior quality to those generated by traditional approaches.

The ability to compare the quality of alternative districtings objectively is therefore a critical element of the research undertaking. It is important to remember that the end results need only satisfy the essential criteria and not necessarily be optimal in order to achieve or exceed the traditional standard. (That is, a satisficing solution is operationally adequate and an optimal solution need not be sought if cost-prohibitive.)

On the other hand, the districting model(s) produced by this research should eventually, if not immediately, be applicable to the entire country in spite of varying settlement patterns and availability of digital data.

They should also provide some fundamental tools for a spatial decision support system that can have substantial practical application in related 'field-based' activities such as taking surveys, establishing polling zones and polling lists, and delivering mail and advertising materials door to door.

1.3 THE IMPORTANCE AND BENEFITS OF THE RESEARCH

In addition to the potential impact on the quality of the Census results, there are a number of probable practical benefits of successfully automating the creation and evaluation processes of collection unit districts:

1. formalizing the districting process and automatically assessing the quality and the operating costs of alternative districtings increase the future possibilities for decentralizing the districting process to the regional offices where the status of existing and planned housing developments is better known;

2. increased automation will reduce the elapsed time needed to produce the districtings which affords opportunities for utilizing more up to-date-information;

3. measures of spatial variability that are needed to automate the decision making

elements (e.g., selecting methods from the "toolkit") are implemented and tested for their relative suitability; and, finally,

4. savings generated from increased automation will help to fortify the cost/benefit ratio of the Census.

1.4 THE STRUCTURE OF THE DISSERTATION

The next chapter establishes the relationship of the research topic to the general districting problem and indicates how elements of the current state-of-the-art can be applied and/or extended to permit the structuring and resolution of the problem.

Chapter 3 defines the specific problem being addressed, describes the context in which it is being solved and indicates the extent to which a solution procedure will be applied and tested.

The proposed solution procedure is detailed in Chapter 4 together with a discussion of the evaluation criteria and methodology for measuring quality.

The scope and results of testing are documented in Chapter 5 together with a discussion of the limitations of the testing and the capabilities.

Conclusions and recommendations for future work and research are reported in the final chapter.

CHAPTER 2

THE CENSUS COLLECTION UNIT DISTRICTING PROBLEM

Districting or "region building" is the process of subdividing a large region into sub-regions or districts of manageable and easily recognizable extent. While one would expect to find large numbers of useful publications on this fundamental operation in geography, the authors of one of the better treatments of this topic have observed that:

> "although regions are traditionally a central theme in geographical writing, geographers have always been curiously reticent about [describing] the ways in which regions can be built up." [Haggett, **et al**, 1977, page 450]

The purpose of this chapter is to define the general districting problem and the particulars of the census collection unit districting problem and to place both in the context of the literature on "region-building".

2.1 OVERVIEW OF THE GENERAL DISTRICTING PROBLEM

This section discusses both the scope and the depth of the general districting problem and relates various elements of the problem to treatments of the districting topic in the literature. The scope is shown through sample works of semi-automated districting applications which are, to some degree, related to the collection unit districting problem. The depth of the previous research in automated districting is indicated through a discussion of alternative districting criteria and by selective reference to studies of the optimality and computability aspects of districting or region-building.

> "Region-building is one of the commonest applied problems encountered in locational analysis. For both the private and public sectors, efficient regional divisions provide one of the ways of reducing the cost of spatial interaction (whether measured in terms of pupils' journey to school, patient flows to hospitals, or efficient marketing movements for a company). Regional divisions represent a compromise between spatial contiguity on the one hand and grouping counties with like characteristics on the other .... The number of possible regional divisions or combinations in any study area is usually very large indeed. Thus, any proposed scheme is less likely to be a single sharply peaked optimum, than one of a set of rather similar near-optimal solutions." [Haggett, **et al**, 1977, page 490]

From a survey of the literature [Massam and Goodchild, 1971; Abler, **et al**, 1972; Helbig et al, 1972; Massam, 1975; Sammons, 1978; Cameron, 1984; Sutcliffe and Board, 1986; etc.] some of the more common constraints and criteria for assessing the level of optimality of a given districting include:

1. Completeness Constraint

All of the entities in the original space must be allocated or classified into one of the resulting district subspaces.

2. Contiguity Constraint

Each member of any given subspace must be adjacent to at least one other member of the same subspace. The resulting partitioning is termed 'spatially constrained'. The decision to join two spatial units is based both on the statistical similarity or complementarity and the spatial contiguity of the two units.

3. Similarity Conditions

Members of a given subspace are required to be 'in some sense' homogeneous. Examples include population characteristics - such as language, age, and income; industrial characteristics - such as agricultural, manufacturing or service activities; or land use types - such as commercial, institutional or residential. Such regions are considered relatively 'uniform'. "Uniform regions may be defined as that arrangement of regional boundaries for which:

$$\frac{\text{External (between region) variation}}{\text{Internal (within region) variation}} = \text{maximum}$$

(cf. analysis of variance).....Uniform regions are sometimes termed homogeneous or formal regions." [Haggett, **et al**, 1977, page 453]

4. Equity Conditions

The districting function in this case is required to distribute the space (i.e., land surface area) or the constituent entities (e.g., persons, households, etc.) in a predefined - usually equitable - fashion. A common equity condition is the requirement that districts distribute a total workload evenly among a fixed size workforce. (See Gaile, [1984] for a review of measures of spatial equity.)

5. Efficiency (or Distance Minimization/Separation) Conditions

The efficiency condition and/or minimizing the distances to be travelled or the separation between elements of a set is a very well researched topic. It is a common approach in the region building in the geographic literature (see Haggett, **et al**, 1977 for a good summary of this approach relative to others). The location/allocation literature spans both geography and operations research and often incorporates a efficiency condition as the basis for optimization (see Hodgart, 1985 and Ghosh and Rushton, 1987 for reviews of various ways in which this can be done). For a taxonomy of partitioning problems for which efficiency conditions such as distance minimization are usually applied see Garey and Johnson, 1978.

Significant economic costs are often directly attributable to the inefficiency of existing districtings. Therefore, a common condition for districting processes is to produce a cost-efficient set of districts. In certain cases, such as the creation of ambulance districts, these savings can be measured in terms of lives saved.

The districting function in this case is expected to partition the space such that the separation between units comprising a subspace is minimized. Alternatively, depending upon the application, the distance factor could incorporate a consideration of the total road distance contained within a subspace as part of a companion equity condition. Using the Pythagorean theorem, distances can be calculated in multidimensional space (i.e., involving several factors) provided that the vectors (i.e., axes) are all orthogonal and thus concepts such as temporal and social distances can be incorporated. It must be stressed that:

> "since the relationship only holds when the individual vectors are orthogonal, we cannot simply use the original variates describing the data since these are likely to be intercorrelated and thus non-orthogonal. Principal components analysis is usually adopted to change the original non-orthogonal vectors into orthogonal components although, alternatively, a generalized distance measure such as Mahalanobis's d^2 statistic could be used to allow for the intercorrelation among the variates. Most grouping programs proceed therefore, not on the original observations but on the component scores (e.g., Steiner, 1965)." [Haggett, **et al**, 1977, page 468]

(See Morrill and Symons, [1977] for an analysis of trade offs between equity and efficiency conditions.)

6. Centrality (Location) Conditions

One form of centrality condition is where the geographic centres of each of the districts are coincident. The resulting pattern is one where the regions are nested and concentric. These types of districts are useful for monitoring continuous phenomena such as radio transmissions, pollution emissions and innovation diffusion from a single central location.

Discontinuous phenomena such as letter carrier routes, or police patrol zones also tend to emanate from a central location such as a post office or police station. In this case, the districting function is required to partition the space such that the variance on the commuting times/distances from the centre(s) is minimized so as to balance the distances to the start of the routes. When the districts are being canvassed or covered from a single central point such as a radiation source or a base camp for search and rescue operations, it may well make sense to use districts based on sectors radiating from the source to minimize distance to the start of the route.

7. Consistency (Preserving Existing Boundaries) Conditions

Helbig has interpreted the consistency condition as the desirability that "districts cross predefined or predetermined boundaries (e.g., rivers, major highways, county lines, etc.) as infrequently as possible ... maintenance of these boundaries simplifies the mechanics ... [and] tends to reduce the number of alternative solutions produced by an algorithm although it could preclude solutions that are potentially superior in terms of ... other criteria" [Helbig, **et al**, 1972, p. 736]. Alternately, it can also be interpreted as including the desire that existing boundaries be preserved for the purpose of facilitating longitudinal analysis (i.e., continuity or stability conditions).

The derivation and implementation of suitable districting functions varies in degree of difficulty in relation to the number of geographic entities and the homogeneity of their spatial distribution - whether expressed in terms of counts or densities. All of these constraints or optimization criteria - **completeness, contiguity, similarity, equity, efficiency, centrality, or preserving existing boundaries (or consistency and continuity)** - should be achievable by the components of a flexible tool kit or model. Further, all of these criteria have some level of relevance to the districting of census collection units. The relevance of each of these criteria are discussed in greater detail in Chapter 3. Finally, the tools should also be able to be used with different general approaches to districting.

2.2 GENERAL APPROACHES TO DISTRICTING

Two approaches to districting are common and may be combined:

1. Disaggregation

Disaggregation is a 'top down' partitioning of a space into zones based on an **a priori** model or geometric pattern. A classical example is the use of a regular grid pattern for searching for hidden treasure if the space is undifferentiated (e.g., a homogeneous plane). The resulting pattern is made up of similar uniform zones that are equitable if the probability of finding the hidden treasure is also uniform and continuous.

For a non-continuous or differentiated surface (made up of 'building block' areas), the disaggregation approach to districting assigns the 'building blocks' to zones on the basis of values of characteristics that refer to higher order spatial units (e.g., grid cells or administrative or statistical units). For example, the allocation of 'building blocks' might be based on the location of the 'building block' (or more specifically its boundary or centroid) relative to the limits of one or more of the higher order units. This is in contrast to an assignment based on some characteristic value associated with the 'building block' (e.g., the number of dwellings, its size or the amount of work it represents). Thus, the disaggregation approach tends to be most effective when the surface is undifferentiated and the phenomena under examination is continuous over the entire space.

This type of districting is termed 'logical division' or 'classification from above' and is identified as "a deductive one [since] we must have prior information on the property being used as an indicator" [Haggett, **et al**, 1977, page 455].

Disaggregation can be straight forward (as for the classical treasure hunt example) if a single 'coverage' or digital map is being overlaid on a given region. Serious technical complications can arise, however, when two or more 'coverages' are overlaid. The major problem with a multiple overlay approach is that it is very difficult to control the size or number of subregions so formed. Consider the "line weave" [Gelinas, **et al**, 1988] effect of overlaying different regionalizations where each was made to conform to common physical features such as a river whose course varies significantly through time. This "line weave" type of problem is often avoidable by employing **a priori** models which can be treated as having fuzzy boundaries. For example, if using **polygon overlay** would generate numerous small and spurious zones along the edges of a near identical set of boundaries, switching to a lower-resolution (i.e., fuzzier) assignment process such as **point-in-polygon** will result in the categorical assignment of each 'building block' unit. Combining the categorized sets of 'building blocks' to form their outer limits will then generate a single, discrete coverage. (See Robinson, [1984] for a discussion of design issues in the application of geographic information systems under conditions of inexactness.)

2. Aggregation

Aggregation is a 'bottom-up' partitioning of space based on decisions to combine individual 'building blocks' based on specific values associated with them. This approach requires that the surface be differentiated and is particularly recommended when the phenomena being studied is not continuous.

In this approach, the characteristics (e.g., position, counts, densities, etc.) relate to the 'building blocks' themselves. "Grouping or 'classification from below' proceeds by grouping individual elements ... according to certain criteria of similarity", [Haggett, **et al**, 1977, page 456]. For example, the standard 'dissolve' function of a Geographic Information System (GIS) forms aggregated units by eliminating shared sides between two zones with a common classification value (e.g., poor, rich, average income, etc.). Equally, the grouping can be based on notions of efficiency as represented in the **assignment model** which iteratively joins 'building block' units on the basis of the proximity of their centroids. New centroids are calculated for the combined zones and the process iterates until the desired number of districts is attained. That is, the aggregation process can be iterative and hierarchical and "the sets so formed may be united to form super-sets, and the procedure can continue until the last union of sets yields the universal set" [Harvey, 1969a, page 330]. Problems are few

> "where a small number of [for example] counties have to be assigned to a fixed number of regions [as] complete enumeration of all possible allocations of counties to regions may be feasible and the 'best' grouping for the purpose at hand chosen .... [However,] Cliff and Haggett (1970) show that the number of alternative combinations rise explosively with n, the number of counties, so this approach is not feasible in most practical situations" [Haggett, **et al**, 1977, page 470].

Indeed, work by Garfinkel and Nemhauser in 1970 [Garfinkel, **et al**, 1970] and recently supported by Goodchild and Hosage shows that $n = 40$ represents "a rough upper limit to the size of the problem which can be handled by branch and bound methods in reasonable CPU time" [Goodchild and Hosage, 1983, page 10].

In the day to day operations of agencies responsible for field activities such as census and survey taking, mail and flyer delivery, reading meters and revising electoral rolls, such 'districts' are created, utilized, and revised on an ongoing basis. This is due, in no small part, to the ability of the human districting specialists to internalize

> "heuristic methods ... which allow (a) isolation of suitable groups of counties to act as 'cores' around which others might be aggregated and (b) exploration of the likely compromises that may be needed to be made in order to attain certain desired mosaic patterns" [Haggett, **et al**, 1977, page 460].

3. Combined

Thus, while districting can be accomplished in essentially one of two ways: disaggregation, and aggregation.

> "In practice, the sharp division between the two main procedures, partitioning and grouping, may not be maintained ... [and] an iterative approach, alternating the two main procedures of division and grouping is likely to offer a useful compromise strategy in region building". [Haggett, **et al**, 1977, page 456]

In most cases where the number of entities to be districted is large, heuristic methods may be necessary to keep the number of computations within reasonable limits without oversimplifying the model unnecessarily.

Therefore, it can be expected that a tool kit/model appropriate for the computerization of the collection unit districting problem will be **multi-stage** and **multi-component**, combine both **disaggregation and aggregation approaches** and rely heavily on **heuristic** tessellation methods that in some sense mimic the strategies employed by districting specialists.

2.3 COMPUTER-ASSISTED COLLECTION UNIT DISTRICTING

Computers have been applied to the creation of functional districts from standardized units such as census blocks, enumeration areas, counties, grid square cells, network links or population centroids for over twenty five years.

Some of the better known early applications include:

1. optimizing school districts [Yeates, 1963; Honey and Kohler, 1978; and Sutcliffe and Board, 1986];
2. optimizing electoral districts [Weaver and Hess, 1963; Morrill, 1973; and Thompson, 1982];

3. optimizing safety service zones (e.g., fire, ambulance, police), [Godlund, 1961];
4. optimizing the linkage between freeway construction and medical care areas [Moore, 1959];
5. optimizing administrative service areas [Massam and Burghardt, 1968; and Massam and Goodchild, 1971]; and
6. optimizing the location of recreational areas [Duffield and Coppock, 1965].

More recently, the range of applications has spread to:

1. radio and cellular telephone repeater station areas [Glick, 1990];
2. sales districts [Glick, 1990];
3. urban level electoral redistricting [Nicholson, 1982; and Van Est, **et al**, 1983]; and
4. developing "environomic units" [Gelinas and van Wyngaarden, 1986].

Most of these problems have been approached 'from below' using aggregation techniques on discrete surfaces. The aggregation methods that have been developed for such applications have employed optimization criteria such as equal-population, equal-areas, equal-density, compactness, or homogenous best fit (constrained or unconstrained) and are based on the minimization of some measure of error or deviation [Moellering and Tobler, 1972].

Elements of each of these alternative criteria can be incorporated and/or adapted to assist with the construction of the major components of the census collection unit districting problem. Each of the aforementioned aggregation methods are well documented in the literature and are briefly described here:

1. Equal-population zoning systems

Sammons [Sammons, 1976] devised a zoning system which minimizes the sum of the absolute deviations from the average population for a given number of zones which used the following objective function.

FORMULA:

minimize $F(G) = \Sigma_k^m \mid \Sigma_{i \epsilon A(k)} \; p_i - \Sigma_i^n (p_i \;/\; m) \mid$

where: m is the desired number of zones;

n is the number of building block units;

G is a set containing the classification of n building blocks into m zones;

$A(k)$ is that subset of G which contains the indices of the building blocks allocated to the kth zone; and

p_i is the population of the ith building block.

2. Equal-area zoning systems

Approximately equal-area regions can be created using Sammons' formula by substituting the population variables in his equation by area variables.

3. Equal-density zoning systems

Similarly, regions with approximately equal population densities can be determined by replacing the population variables in Sammons' equation with density variables.

4. Compact zoning systems

Approximately compact zoning systems can be produced by maximizing some measure of zone shape, usually formulated in relation to the shape of a circle with equivalent area. Early efforts at political redistricting emphasized the importance of this criterion because of its relationship to possible attempts at gerrymandering [Reock, 1961]. Consequently, significant effort has been devoted to effective measures of compactness [Bunge, 1962; Lee, **et al**, 1970; and Kimerling, **et al**, 1971]. Massam and Goodchild have developed an index of shape for non-homogeneous surfaces that is based on "the moment of inertia of a *disc* around its centre of gravity [in proportion to] ... the moment of inertia of an *administrative unit* around its centre of gravity [such that] ... both the disc and the administrative unit are the same area" [Massam, 1972, p. 4].

In cases where gerrymandering is not of concern, it is not immediately clear that there is a direct relationship between the shape of the districts and the equity of the workload (especially when expressed in terms that combine the number of dwellings to be visited with the distances to be traversed).

5. Homogeneous or "best fit" zoning systems

Cliff and Ord [1975, pages 17 - 19] suggest that an ideal aggregation procedure preserves as much of the original variation of the variable as is possible. Openshaw [1978, page 789] recommends an objective function that:

1. maximizes the degree of intra-zone homogeneity,
2. maximizes the variation of the independent variable to yield an unbiased estimate of 'b', (which is the coefficient of the independent variable and is an ordinary least squares - OLS - estimate)
3. maximizes the variation in the independent variable relative to that of the dependent variable. This function is based on [Blalock, 1964] and [Hannan, 1971] and can be expressed formally as:

$$\text{maximize } F(G) = \frac{\Sigma_k^m \Sigma_{i \in A(k)} (y_i - y_k)^2}{\Sigma_i^n (y_i - \Sigma_i^n y_i / n)^2} - \frac{\Sigma_k^m \Sigma_{i \in A(k)} (x_i - x_k)^2}{\Sigma_i^n (x_i - \Sigma_i^n x_i / n)^2}$$

This relationship can also be expressed in terms of optimizing the correlation between the independent and dependent variables (x and y) aggregated to a given number of zones to obtain a best-fit model zoning pattern.

Williams [1976], in order to improve the precision of regression estimators derived from grouped data, has suggested a method that minimizes the sum of the parameter standard errors. Since the researcher usually wants the magnitude of the standard errors to be as small as possible in relation to the size of the parameters, this criterion is reformulated as a sum of the ratios of the standard errors to parameter sizes [after Openshaw, 1978, page 788].

However, it is often necessary to trade-off "goodness of fit" of a model for constraints on the design of zones that are also relevant to the model. These trade-offs or constraints can be expressed as assumptions relating to the linear-regression model and the estimates of the OLS parameter. Some of these constraints include:

1. ensuring that the mean residual is zero;
2. ensuring zero spatial autocorrelation among the residuals;
3. ensuring that the independent variable is not autocorrelated after aggregation; and
4. ensuring residual homoscedasticity (based on Spearman's correlation coefficient) between the absolute values of the residuals and the independent variable [Mather, 1976, page 80 - 81].

Openshaw has concluded that "only a random zone-design procedure will produce unbiased estimates of b" and that "there is going to be no simple or general-purpose solution to the problem" [Openshaw, 1978, page 793].

In general, computer-assisted implementations of districting functions that are heavily weighted by absolute and relative locational constraints are particularly difficult to construct. On the other hand, manual methods are seldom cost/effective whenever the number of entities to be manipulated becomes large and/or the level of complexity of the districting function increases.

Disaggregation approaches tend to be used for undifferentiated or continuous surfaces for applications such as search and rescue, forest fire fighting and disaster evacuation planning. This typically employs tessellations of continuous surfaces using regular structures such as rectangular, circular (i.e., sectors and rings) or other polygonal grids (based on triangles, hexagons, etc.) with uniform cell sizes in terms of area covered.

Regular geometric tessellations can also be adapted to account for variable distributions over a space comprised of discrete elements. This is an aggregation approach and is analogous to the use of extensible cells [Tamminen, 1984] for efficient geographic data storage and retrieval or the use of irregular rectangular grids cell tessellations for reducing the number of cells needed to represent a surface with highly varying topography [Makarovic, 1979].

As depicted in Figure 2.1, these rectangular grids can be unidirectional or bi-directional and can have regular shapes of uniform area (disaggregation) or have irregular shapes resulting in a relatively uniform distribution of discrete elements between cells (i.e., aggregation). Minor difficulties must be overcome in the case of bi-directional grids if the desired number of cells is a prime number since the number of rows and columns will not be uniform over the entire space.

It has been demonstrated by Tobler, [1973] that an alternative methodology is to transform the shapes of the building block units to give a cartogram of uniform density and then apply a regular grid to partition that space into equitable units. Reversing the transformation yields district boundaries on the original cartographic reference base. This approach was not pursued, however, due to expected high costs of operation.

Regular ring and/or sector geometric models are useful for tessellating a continuous surface if the phenomena under study can in some way be related to a centre or focal point. One common example is to tessellate the spatial distribution of a radio or television audience about a central point like a broadcast transmitter. Sample representations of unidirectional (rings or sectors) and bi-directional (rings with sectors -- although sectors with rings are also feasible) are depicted in Figure 2.2. Again, the case of a prime number of cells requires special attention.

In both cases, rectangular and circular grids, a change in the choice of origin and orientation will provide different partitioning of the given space, as does a change in the ratio of rows to columns or rings to sectors. The geometric models must be related to the shape of the given space (either by the bounding box which is based on minimum and maximum coordinate values or by the minimum circumscribed circles can be used to establish a default origin and, if needed, an orientation for the grid). How this is done will also influence the location of the individual districts.

Finally, a discrete surface can be tessellated based on the proximity of the discrete elements (usually represented by their centroids) to a supplied set of seeds for disaggregation (or on the basis of both proximity and equity of allocations in the aggregation case). Examples of regular and random seed locations are provided in Figure 2.3. Here, there is virtually no difference in the tessellation methods for prime and non-prime numbers of cells.

Figure 2.1 Rectangular Grid Tessellations

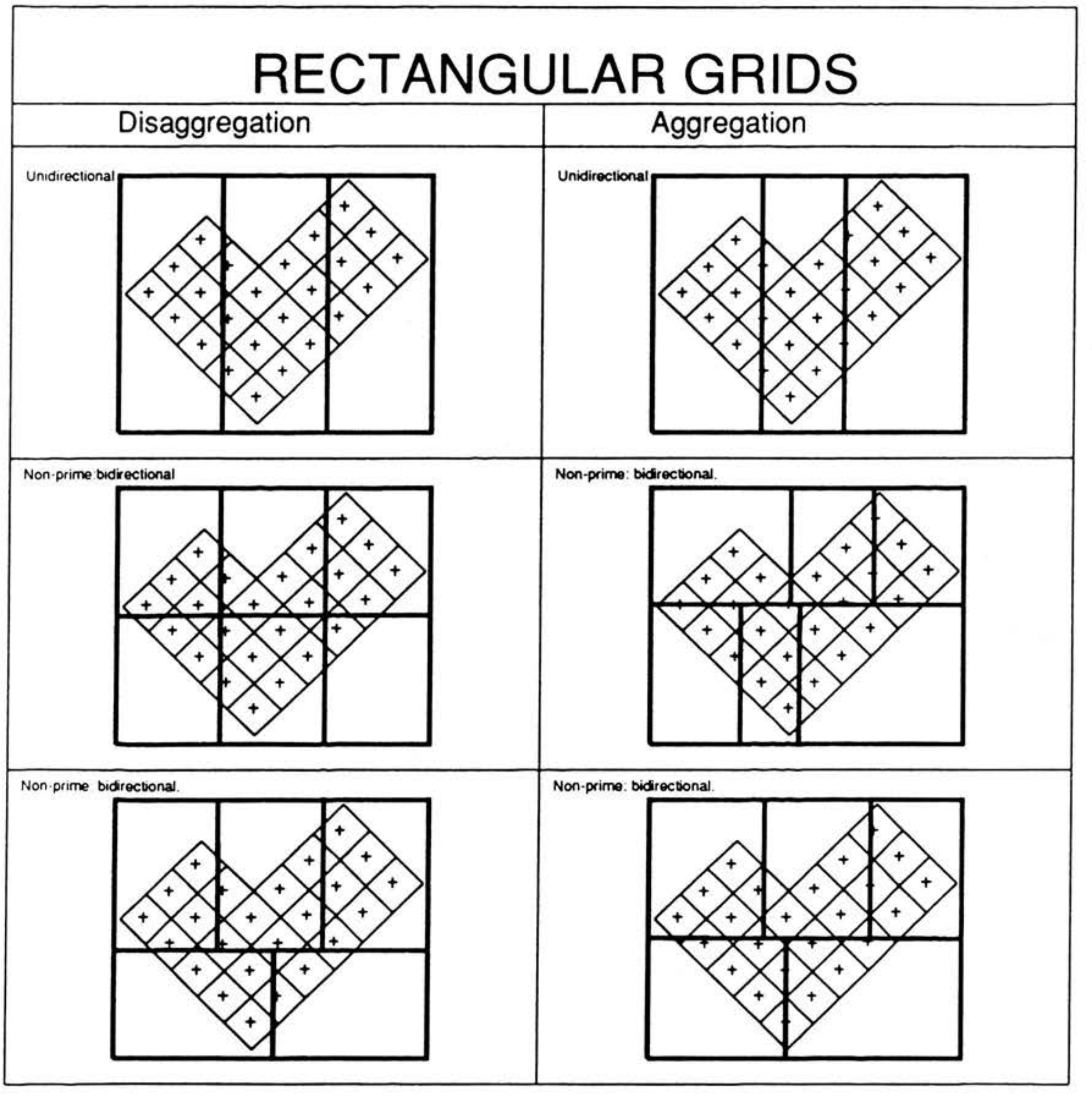


Figure 2.2 Circular Grid Tessellations

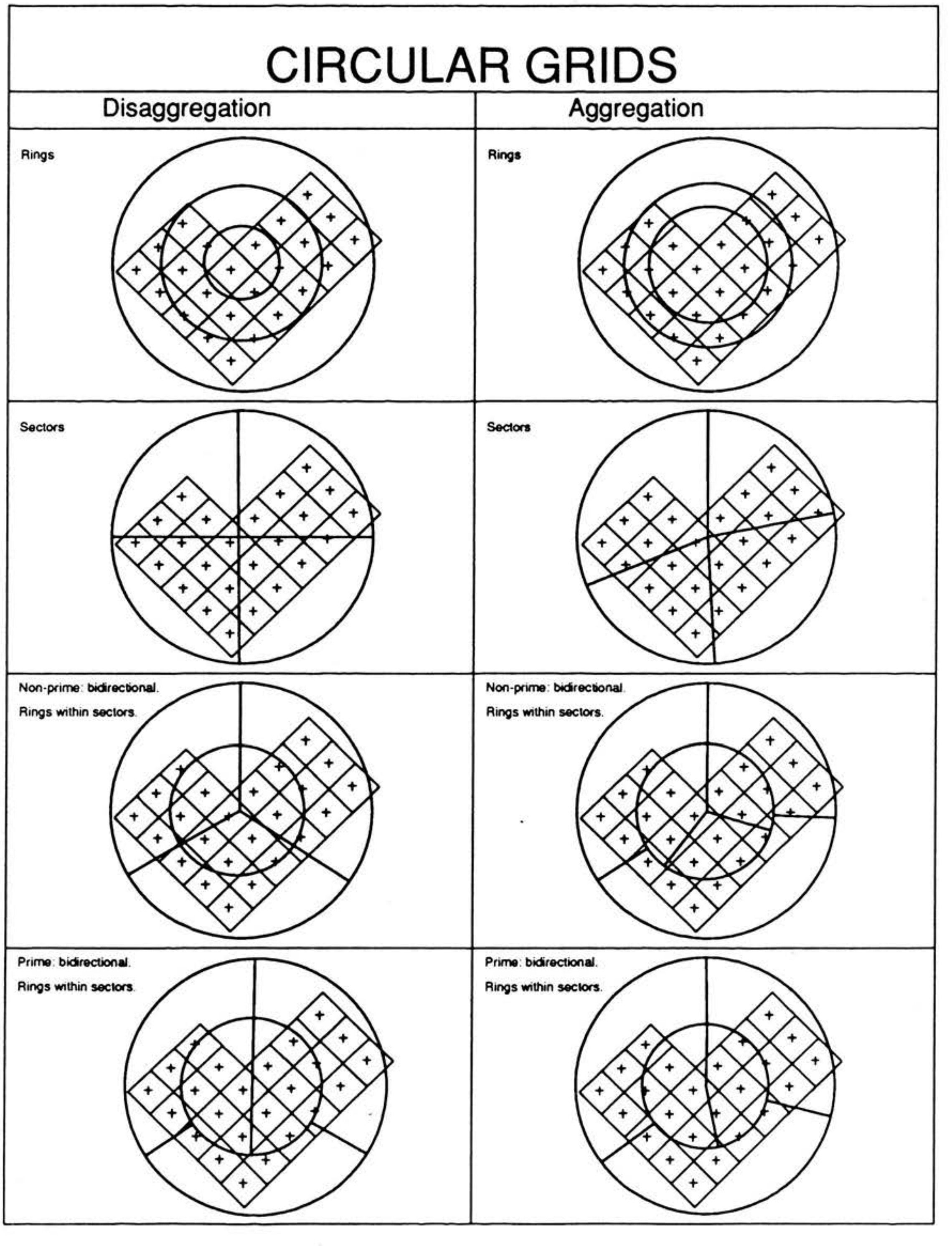
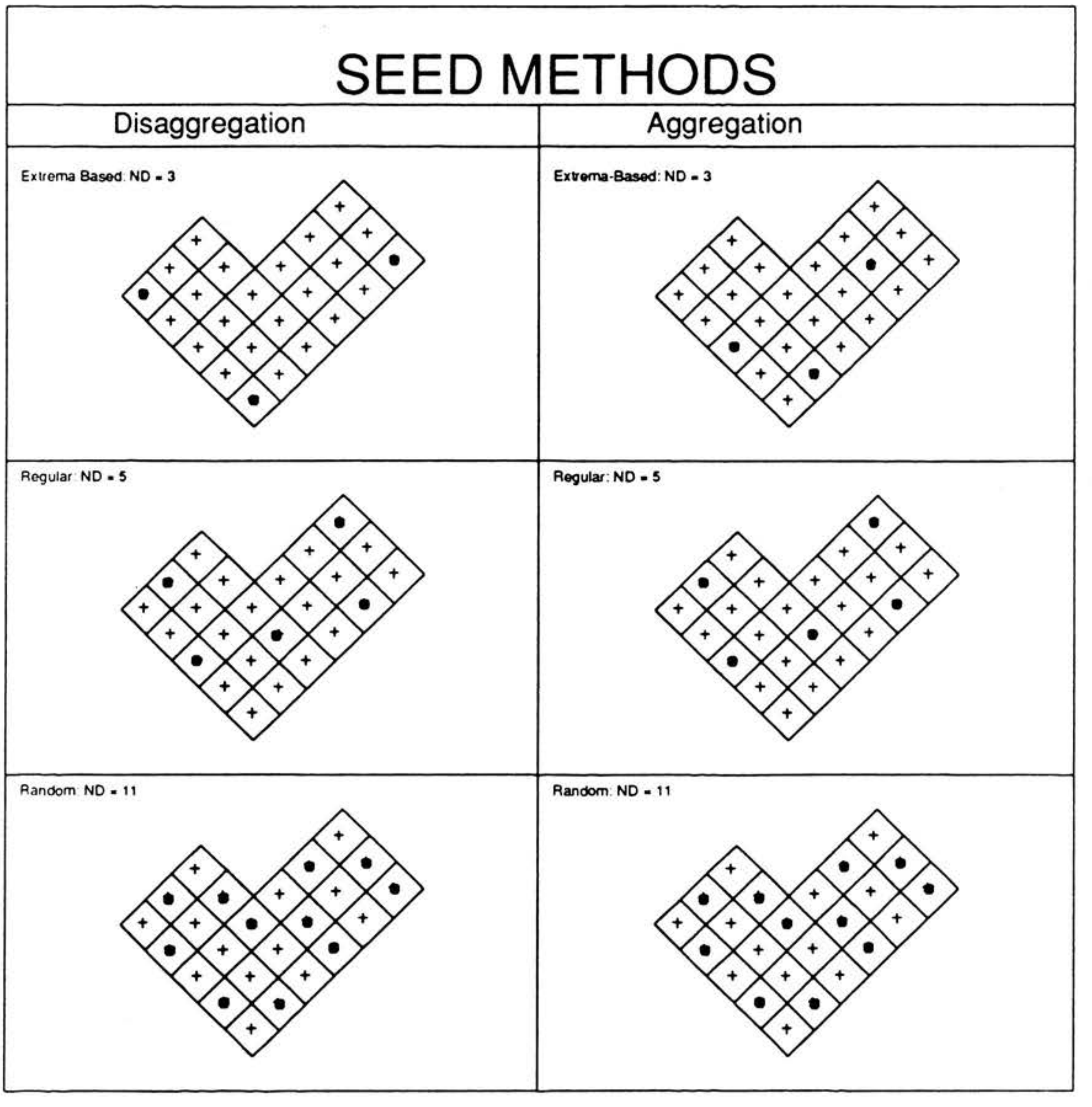


Figure 2.3 Seed Based Tessellations



These classical geometric models (that are easy to visualize and comprehend by operations staff) are useful elements of a districting "tool kit". In addition, a third method of seed selection has been devised as part of this research effort and is included in the districting "tool kit" and is termed 'extrema-based seeds'. This algorithm identifies as seeds either the centroids which (as a set) are most separate in terms of their location, or the centroids which have the highest values for a particular attribute.

It is unreasonable to expect such straight-forward models to generate tessellations which are able to meet a large and varied set of criteria specific to the districting of Census collection units in all cases. It is likely, however, that operationally adequate districts can be generated at low cost in a significant percentage of cases. Therefore, a framework for incorporating these geometric models into a "tool kit" together with other useful and original districting methods is described in Chapter 4. This description of the tool kit is preceded by a review of other existing computer-assisted approaches to the census collection unit districting problem (in the next section) and the detailed specification of the collection unit districting problem in Canada (in the next Chapter).

2.4 EXISTING SYSTEMS FOR CENSUS COLLECTION UNIT DISTRICTING

The only identifiable implementation of computer-aided capacity for creating Census collection units is a prototype system developed at the U.S. Bureau of the Census. The earliest results from a prototype system for the 1980 Census that operated on-line were reported in 1980. For a given region, "blocks were automatically accumulated in block number order until 400 households were reached" [Bonnette, 1980, page 110]. The proposed grouping was then accepted or modified interactively by the operator who identified blocks that should "be omitted or included in the [Enumeration District or] ED." [ibid, page 110]

Only household counts and adjacency information inherent in the manually-generated block numbering system were employed in this system and no attempt was made to determine optimality nor to measure efficiency. The resulting system was considered experimental and was not operationally employed for the 1980 U.S. Census "primarily due to the unavailability of the digitized [Geographic Base Files or] GBF's." [ibid, page 110]

The U.S. Bureau of the Census is extending its coverage of cartographic files (formerly called "DIME files" and now improved and called "TIGER files") to the entire country to serve as a basis for both collection unit creation and collection mapping for the 1990 Census. Further, they are reported [Marx, 1985] to have enhanced the prototype system design to consider "block adjacency rather than numeric block sequence" as "the primary factor in the determination of the 'cluster'". But, they did not use computer assisted approaches for the creation of collection unit districts, termed 'address register areas' or ARAs, for the 1990 Census [Marx, 1987]. Instead, the groups of blocks forming the ARAs were established by field representatives on a distributed and manual basis.

In Canada, on the other hand, the prototype computer-assisted districting system, developed for this dissertation, has been converted into a production system for use on a majority of large urban centres for the 1991 Census of Canada.

2.5 CONCLUSIONS

After a review of the literature, including some of the literature on location-allocation modelling [Lea, 1973; Hodgart, 1985] and on political redistricting [Goehlert, 1981], it is evident that the various approaches to the general districting problem have been well summarized by Haggett, Cliff and Frey in 1977 and that there has been relatively little work of direct relevance to the Census collection unit districting problem in the intervening period. It is equally evident that the specific districting application under study in this dissertation has yet to receive much attention in the literature beyond the 1980 paper by a staff member of the U.S. Bureau of the Census.

The next chapter defines the collection unit districting problem in the specific context of the census taking activities in Canada and has, therefore, a much less theoretical orientation. It also delimits the geographical extent for which the testing of the research results have been conducted.

CHAPTER 3

CENSUS COLLECTION UNIT DISTRICTING IN CANADA

3.1 PURPOSE OF THIS CHAPTER

This chapter considers the various dimensions of the Census collection unit districting problem in Canada. It shows that a relatively large number of constraints must be considered during the districting process. Further, it indicates that there are few, if any, formal rules for making choices or trade-offs. That is, there are no quantitative measures of quality in the current manual districting process.

For those unfamiliar with the various spatial frameworks and related terminology of the Canadian Census, this chapter begins by providing some background into those structures and terms.

3.2 BACKGROUND

The major reasons for taking the Canadian Census are related to the practical needs of government. Indeed, each Census the proposed Census questions and sampling ratios are approved by Cabinet. The British North America Act (1867), establishing the Dominion of Canada, mandated the taking of a census every ten years, starting in 1871, to provide the framework for the redistribution of the population into federal electoral districts containing approximately the same population for equal representation in the House of Commons. This remains as the single most important reason for taking decennial censuses in Canada.

The need to monitor the effectiveness of federal government programs and to determine the distribution of federal funds provided on a per capita basis has led to the implementation of a quinquennial Census in years ending in six (6) since 1956.

To provide a basis on which to organize the taking of the census, every five years, Statistics Canada partitions the entire country into approximately 40,000 units. Re-districting of Census collection units is needed because of changes in population distributions and changes to the boundaries of administrative areas.

These collection-based units often also serve as the smallest unit for the dissemination of Census results. Censuses provide the relatively high resolution data needed to permit academics, planners, policy makers and other researchers to study and understand the changing socio-economic fabric of a country or region. Two concerns of these users of census data at the level of the collection unit are:

1. the stability of the collection unit which is needed to facilitate longitudinal analysis, and

2. the biasing of aggregate statistical information due to the methods by which the boundaries are established.

In Canada, these 'dissemination' concerns are subordinated to 'collection' concerns. This is due to, firstly, economic necessity, and secondly, because alternatives exist (e.g., blockface geocoding) for ensuring that these 'dissemination' needs can be met. That is, the fact that the Census is geocoded to the level of the blockface means that customized aggregations (according to user specified areas) of census data can be tabulated (e.g., by homogeneous neighbourhoods or by constant geographic units).

3.2.1 Collecting Census Data

A census involves taking a complete inventory of the amount and distribution of a given entity or set of entities (e.g., people, dwellings, livestock, crops, or industries) for a prescribed space at a particular time. As such it belongs to the family of monitoring, surveillance and inventory problems.

While there is legislation that can be relied upon to force individuals to supply the required data, in practical terms, the success of the Census very much relies on the goodwill of the respondents, the clarity of the questions, the effectiveness of the enumerators and the appropriateness of the collection methodology.

3.2.2 Collection Methodologies

The methodology for collecting the census varies throughout the country depending upon settlement densities, accessibility, and existing information infrastructures for collection, communication and dissemination (e.g., postal and telecommunication services). The most common methodology for census taking involves 'dropping off' questionnaires at every dwelling and then having the respondents return the completed questionnaires to collection points throughout the country by return mail. The U.S. Bureau of the Census, using commercially prepared mailing lists, carries this process one step further and distributes the questionnaires for a large segment of the urban population via the U.S. Postal Service (i.e., 'mail out/mail back').

At the other end of the spectrum of alternative methodologies is the use of canvassers or 'enumerators' as interviewers to obtain the needed information. In Canada, given the much greater cost of this approach, it is utilized only where repeated follow-up visits by the enumerator (e.g., very remote areas in the far north) for the 'drop-off/mail-back' methodology would be cost-prohibitive.

Approaches developed during this research undertaking have national applicability but, for practical reasons which will be presented later, the scope of the testing of the models is restricted to larger urban population centres in Canada. A typical scenario for census data collection in the urban context is:

1. questionnaire 'drop off';
2. questionnaire 'pick up' (or 'mail back');
3. home edit (e.g., for missing data);
4. telephone follow up; and
5. field follow up.

In addition to ensuring that the canvassing encompasses the complete set of entities, it is essential that a given entity be canvassed only once. Given the large number of census takers operating in the field at the same time it is important that the boundaries of the collection units be clearly and unambiguously identified.

3.2.3 The Canadian Census Collection Unit Districting Problem

Census collection units are called "enumeration areas" (EAs) in Canada. Currently they are districted using labour intensive and highly individualized manual techniques, described in detail in Appendix C. The process involves the centralized delineation of a preliminary set of collection unit districts which are revised on the basis of 'field checks' in selected regions of the country, and which are amended (usually by splitting existing units) as required during the actual taking of the census.

The criteria and methodology for manually creating districts for collecting census data in Canada have been employed and continually revised since the earliest censuses. Since 1956, cartographic representations of the limits of each district, superimposed on a suitable base map, have been provided to the enumerators.

Because a major question addressed by this research is whether or not collection districts of comparable quality can be created using computer-assisted techniques, it is important to both explicitly define the nature of the problem, and to identify and quantify the various characteristics that are considered when comparing the quality of different results.

3.3 THE NATURE OF CENSUS COLLECTION UNIT DISTRICTING

In Canada, the purpose of the collection unit districting process is to create a coverage of the given space such that:

1. the entire space is exhaustively and uniquely covered (i.e., categorical or "one-to-one and onto");
2. the districts respect the hierarchy of the standard geostatistical zones (as depicted in Figure 3.1);
3. the districts are of a manageable size for the data collection task;
4. the limits of the districts are readily identifiable on the ground (i.e., conform to such features as represented in a cartographic data base);
5. the settled portions of the districts are, where possible, readily accessible (e.g., the access routes to islands or remote areas are important to consider in establishing areal assignments);
6. the linguistic composition of the districts is essentially unilingual English or French with the number of bilingual districts (i.e., those districts with more than 10% English and 10% French mother tongue) kept to a minimum to reduce the difficulty of recruiting qualified bilingual staff;
7. the method of data collection is uniform (i.e., mixed modes of collection are prohibited within a given collection unit district);
8. the distribution of work over the set of districts is equitable;
9. the district boundaries are maintained from the previous census whenever possible,
10. the number of farms in a rural district does not exceed 99.

3.3.1 The Dimensions Of The Collection Unit Districting Problem In Canada

Table 3.1 helps to characterize the dimensions of the collection unit districting problem in Canada. Since the focus of testing of the proposed model is in the larger urban centres, each of the relevant aspects is presented as an aggregate for the nation as a whole and for those centres of population with 50,000 or more persons.

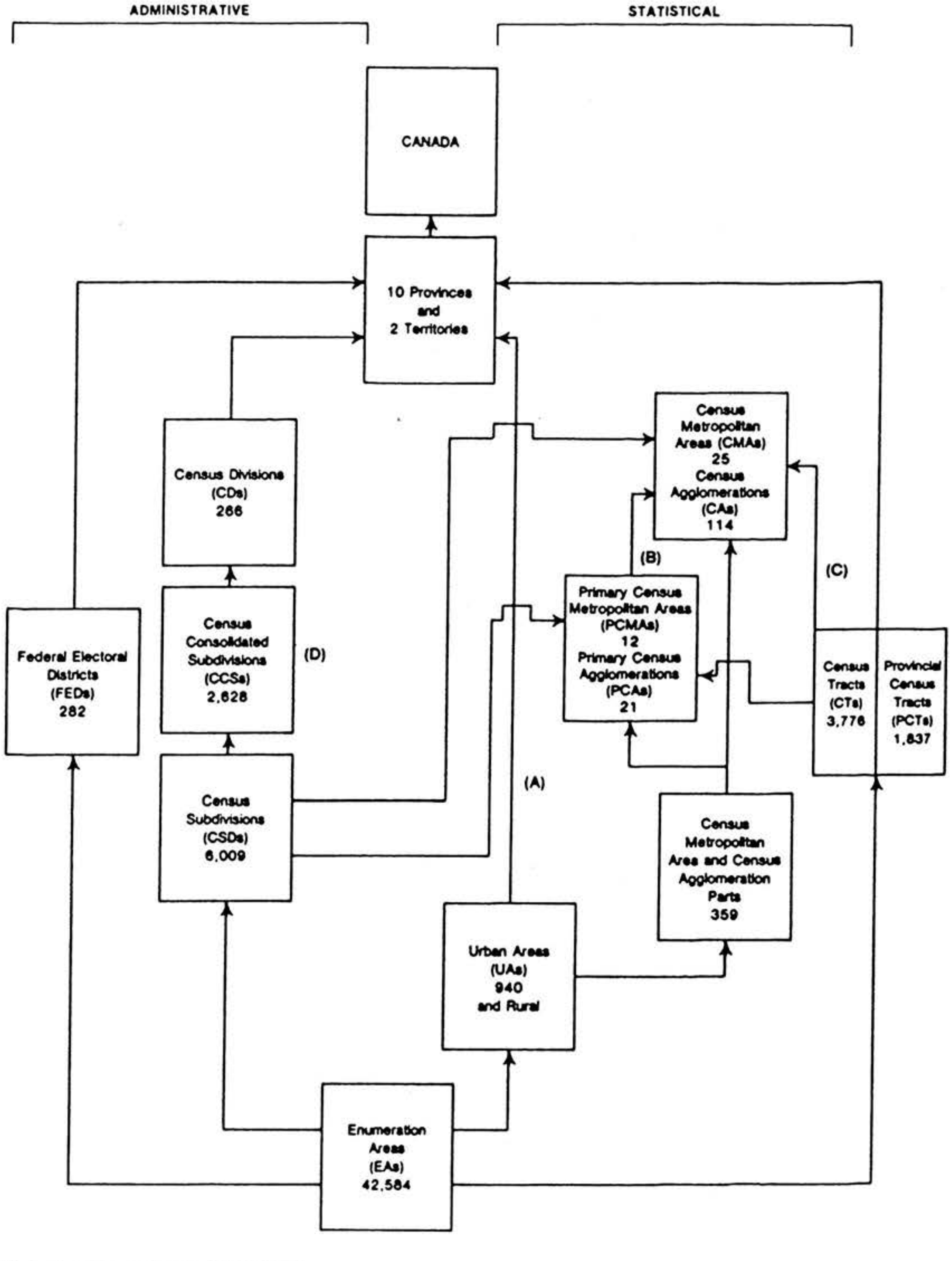
TABLE 3.1 THE SIZE OF THE COLLECTION UNIT DISTRICTING PROBLEM IN CANADA

RELEVANT ASPECT:	LARGER URBAN CENTRES	1986 TOTAL
TOTAL LAND AREA (in square kilometres)	15,277	9,173,000
TOTAL POPULATION	14,571,510	25,354,064
TOTAL NUMBER OF OCCUPIED PRIVATE DWELLINGS	4,800,000*	9,057,533
TOTAL NUMBER OF BLOCK FACES	653,276	2,400,000*
TOTAL NUMBER OF BLOCKS	260,000*	600,000*
TOTAL NUMBER OF ENUMERATION AREAS	20,988	42,584
TOTAL NUMBER OF CENSUS TRACTS	3,776	5,613**
TOTAL NUMBER OF CENSUS SUBDIVISIONS	254	6,009
TOTAL NUMBER OF CENSUS DIVISIONS	7***	266
TOTAL NUMBER OF FEDERAL ELECTORAL DISTRICTS	92***	282

* Estimated values
** Includes 1,837 Provincial Census Tracts (PCTs)
*** Complete Units only

Collection unit districting must support the effective and accurate aggregation, retrieval and tabulation of the Census data according to pre-determined geographical and statistical classification schemes. Figure 3.1 depicts the relationship between the collection units or enumeration areas, and the standard Census geostatistical hierarchy. Descriptions of each of these spatial units are provided in the **1986 Census Dictionary** (Statistics Canada, 1987).

FIGURE 3.1 CANADA'S GEOSTATISTICAL FRAMEWORK



(A) 5 urban areas cross provincial boundaries.
(B) 12 of the 25 CMAs and 2 of the 114 CAs are broken down into PCMAs/PCAs.
(C) All 25 of the CMAs, but only 12 of the 114 CAs, have a census tract program.
(D) Defined by Statistics Canada, in conjunction with the provincial authorities, as a statistical area.

Source: Statistics Canada, 1987, page 121

The districting of over 40,000 census collection units for Canada is a very large, expensive and time consuming task. It must take into account Canada's widely varying dwelling density and the numerous constraints that enable the collection unit to serve its myriad functions.

3.4 CONSTRAINING THE DOMAIN OF THE PROBLEM

As described in the next section, the current manual process for creating EA collection districts in Canada has been adapted to account for varying settlement densities and for the varying resolution of information about the distribution of individual dwellings.

While it would be desirable to test the proposed general purpose model and methodology throughout Canada, for practical reasons, it is currently necessary to focus the application and testing of the model to those areas where all the requisite information (geographic, cartographic, and statistical) is already available in machine processable form. (See Slocum, et al, [1984] to gain an appreciation of the complexity and the difficulty of building a data base suitable for districting.)

In the Canadian context, this means initially constraining the testing of the model to those urban centres of population 50,000 or more for which geocoding of census information at the level of the blockface has taken place and for which Census Tracts have been defined. Table 3.2 lists the areas in Canada where Census Tracts have been created (through consultation with local authorities). Outside of the Census Tracted areas, Census Geography staff unilaterally (in most cases) define a complementary set of units called Provincial Census Tracts (see Appendix D for further details). Figure 3.2 depicts the content of the Area Master Files (AMFs) together with the dwelling count information tabulated at the blockface level.

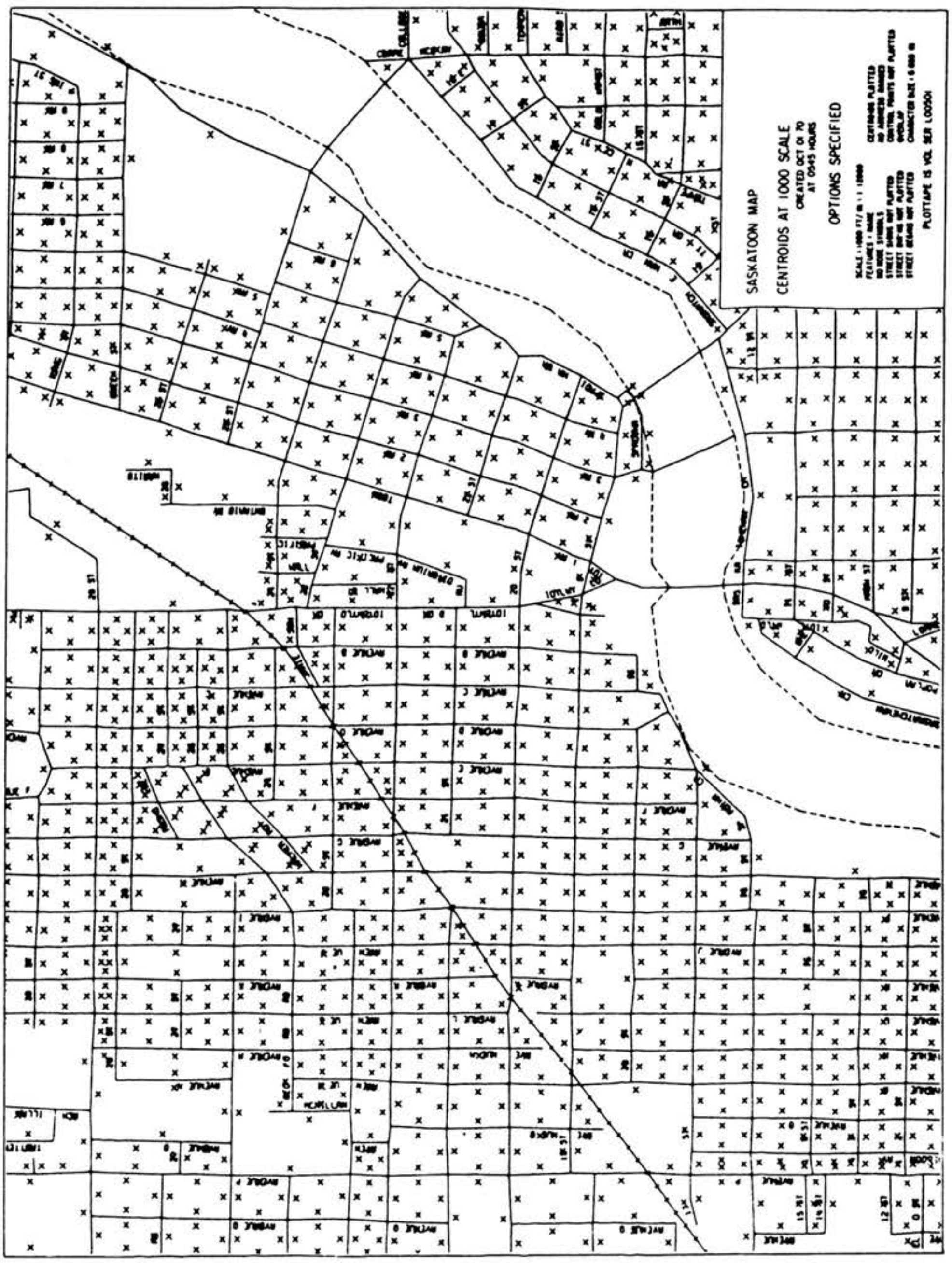
Table 3.2 CENSUS TRACTED AREAS IN CANADA

Brantford (CA), Ontario	Calgary (CMA), Alberta
Chicoutimi-Jonquiere (CMA), Quebec	Edmonton (CMA), Alberta
Guelph (CA), Ontario	Halifax (CMA), Nova Scotia
Hamilton (CMA), Ontario	Kamloops (CA), British Columbia
Kelowna (CA), British Columbia	Kingston (CA), Ontario
Kitchener (CMA), Ontario	Lethbridge (CA), Alberta
London (CMA), Ontario	Moncton (CA), New Brunswick
Montreal (CMA), Quebec	North Bay (CA), Ontario
Oshawa (CMA), Ontario	Ottawa-Hull (CMA), Ontario-Quebec
Peterborough (CA), Ontario	Prince George (CA), British Columbia
Quebec (CMA), Quebec	Regina (CMA), Saskatchewan
Saint John (CMA), New Brunswick	Sarnia (CA), Ontario
Sault Ste. Marie (CA), Ontario	Saskatoon (CMA), Saskatchewan
Sherbrooke (CMA), Quebec	St. Catharines-Niagara (CMA), Ontario
St. John's (CMA), Newfoundland	Sudbury (CMA), Ontario
Thunder Bay (CMA), Ontario	Toronto (CMA), Ontario
Trois-Rivieres (CMA), Quebec	Vancouver (CMA), British Columbia
Victoria (CMA), British Columbia	Windsor (CMA), Ontario
Winnipeg (CMA), Manitoba	

(CMA): Census Metropolitan Area (CA): Census Agglomeration

FIGURE 3.2 AREA MASTER FILE CONTENT

COMPUTER-PLOTTED STREET MAP OF SASKATOON (PORTION)



Source: Statistics Canada, 1971, page 15.

3.5 THE MANUAL METHOD OF DISTRICTING IN CANADA

The process of districting an area into collection units for the Canadian Census is essentially an acquired art. Within the guidelines and constraints listed above, the staff learns (through trial and error) undocumented, subjective strategies to generate a final districting that conforms to the mandatory constraints and guidelines. The data on which these decisions are based are at best variable in currency (up-to-dateness), resolution and quality.

The staff responsible for districting the country into collection units can call upon useful information from a variety of sources:

1. the revised (in the field) individual collection unit maps from the previous Census;

2. the list of individuals and dwellings (by block) contained in the visitation records from the previous census;

3. statistical data from the previous census showing mother tongue, farm, dwelling and population counts by EA;

4. target workloads by collection unit type;

5. dwelling counts (if available) by blockface, retrieved from the previous census data base or compiled from field checks for the Labour Force Survey by staff in the Regional Offices; and

6. lists of apartments compiled for the Labour Force Survey.

Since only very generalized written procedures for either performing the actual districting or for evaluating independently generated results have been documented, the final results of the manual process can only be judged subjectively as either acceptable or unacceptable.

As new information comes available, perhaps from a field check or during the actual field enumeration process, changes are made to the collection unit boundaries to ensure their continued conformance with the constraints and guidelines.

3.5.1 Overview

The following tasks (compiled from written procedures [Seguin, 1976a and 1976b] and through interviews with supervisory staff) comprise the major stages in the manual creation of collection units:

A. Preparation of materials

1. obtain and prepare map manuscripts;
2. prepare a master list from various listings to indicate Census Divisions and Subdivisions, Census Tracts, Provincial Census Tracts, Urban Areas, and dwelling and language counts and forms for recording the correspondance between the new EA and EAs from the previous census;
3. transcribe geostatistical units to map manuscripts;
4. assign the census year and the NTS (National Topographic Survey) scale and index number to the map sheets that have been assembled into manuscripts of a manageable size for districting, measurement, field check and drafting operations, and to facilitate storage and retrieval;
5. delimit and verify the Federal Electoral District boundary and then delimit and colour-code all remaining geostatistical areas; and
6. finally, indicate collective dwellings, and military and penal establishments, that meet a minimal threshold on size to be classified as collectives:

Table 3.3 COLLECTIVE DWELLING SIZE THRESHOLDS (1981 & 1986)

Unit	Threshold Size
Collective Dwellings:	
Hotels, Motels, Tourist Homes	
Lodgings - houses	200 beds
School residences	150 beds
YM/YWCAs, Missions Hostels	200 beds
Work camps	150 beds
Religious institutions	150 beds
Orphanages and children's homes	75 beds
Nursing homes, Old age homes, and	
Chronic Care Institutions	75 beds
General Hospitals	75 beds
Psychiatric Hospitals	75 beds
Juvenile Delinquent homes	75 beds
Corrective and Penal Institutions	75 beds
Hutterite colonies	any
Jails	any
Military camps	any
Other	

B. Districting Process

1. identify constraints;
2. reconcile map manuscripts and visitation records; and
3. create collection unit districts as described at some length in section 3.5.2.

C. Verfication and Correction

1. analyze districting results;
2. process revisions (e.g., municipal limit changes);
3. make modifications and/or improvements; and
4. accept as final.

D. Revision of the Proposed Districting Based on Field Checks

In the general case, dwelling information used in the districting process comes from the previous census which is typically 3 to 4 years out of date at the time it is used. Since much of this information is only available at the level of the previous collection units, it is necessary to conduct pre-census 'field checks' in selected parts of the country. The purpose of these 'field checks' is to confirm that the proposed districts conform to the stated criteria and guidelines. If, for example, major changes have occurred since the previous census, revisions to the district boundaries will be necessary. After the field check has been completed, the collection unit boundaries are 'finalized'.

E. Distribution of Field Collection Documents Based on 'Finalized EAs'

Once the boundary and identification code of a collection unit are finalized, the colour-coded boundary and code information on the map manuscript are manually transcribed to a 'fine drawing' base map for subsequent reproduction. Sufficient numbers of copies are then produced to ensure that each Census Representative is provided with a graphic description of their geographic area of responsibility. These individual collection unit maps together with composite maps for groups of collection units, known as Census Commissioner Districts, are then provided to the Regional Offices who supervise the census collection operation.

F. Post-Census revision of 'finalized' Collection Units that were changed during the census collection process.

In spite of the best efforts of the districting staff and the revision process associated with the field checks, it may be necessary to make changes to 'finalized' collection units in the field during the census collection process.

This kind of change is typically restricted to splitting a given collection unit into two or more new collection units and can lead to sub-optimal work assignments since the parts are too large to be combined into a single workload and yet are to small to be individual workloads themselves and may not be easily combined with other small units nearby.

The main cause of changes to collection unit boundaries is an increase or decrease in settlement densities that can take place in the 8 to 12 months since the districting was finalized. The amount of this change would be much greater were it not for the fact that the municipal boundaries are arbitrarily frozen for statistical reporting purposes by the Census on January 1st of the Census year.

3.5.2 Creation Of Preliminary Collection Unit Districts

A number of criteria govern the collection unit districting process. They include:

1. **Conformance with the Geostatistical Hierarchy**

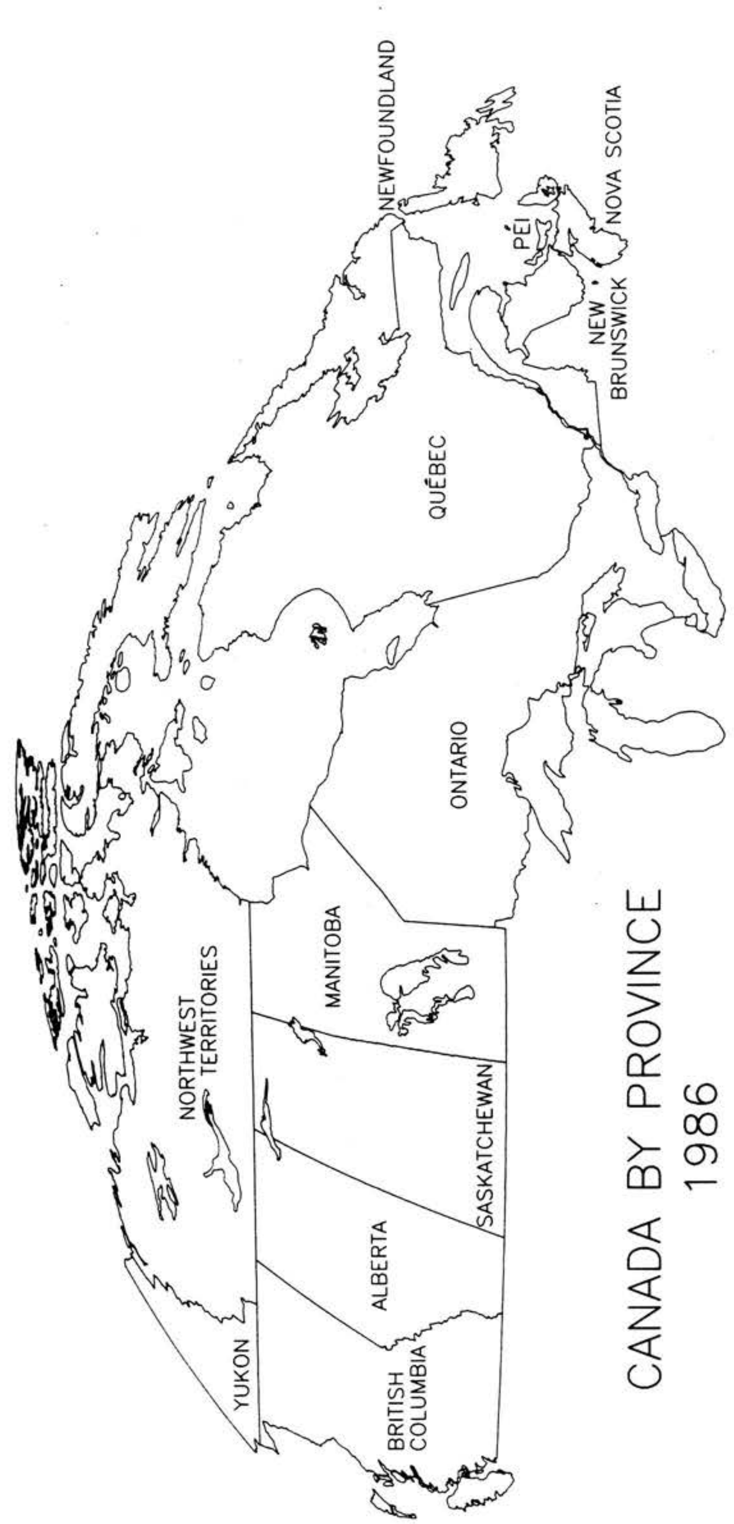
The limits of the collection units must respect, at all times, the boundaries of all other geostatistical units in the hierarchy. Table 3.4 lists the units involved and their numbers.

Table 3.4 HIGHER-ORDER GEOSTATISTICAL AREAS (1986)

GEOSTATISTICAL AREA	URBAN	RURAL	TOTAL
Provinces and Territories	-	-	12
Federal Electoral District	92	190	282
Census Division (County)	7	259	266
Census Sub-Division (Municipality)	254	5755	6009
Census Tracts	3776	1837	5613

Figures 3.3, 3.4, and 3.5 show the higher order geostatistical areas that set the context for the selection of a test area. This selection is described further in Chapter 5.

FIGURE 3.3 PROVINCES AND TERRITORIES OF CANADA



Statistics Canada, CARTLIB

Produced by Statistics Canada, 1987.

FIGURE 3.4 PROVINCES AND CENSUS DIVISIONS OF CANADA

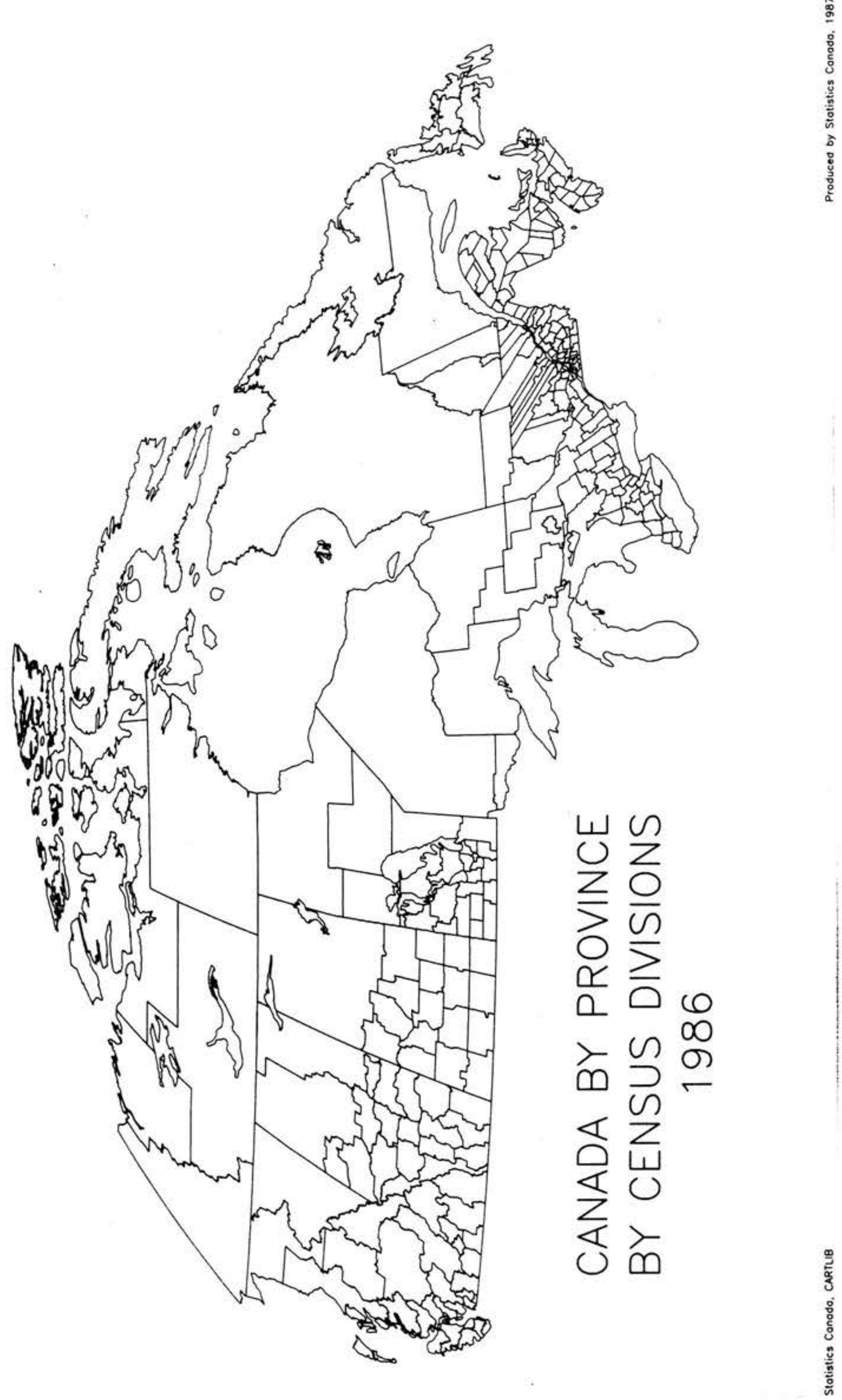
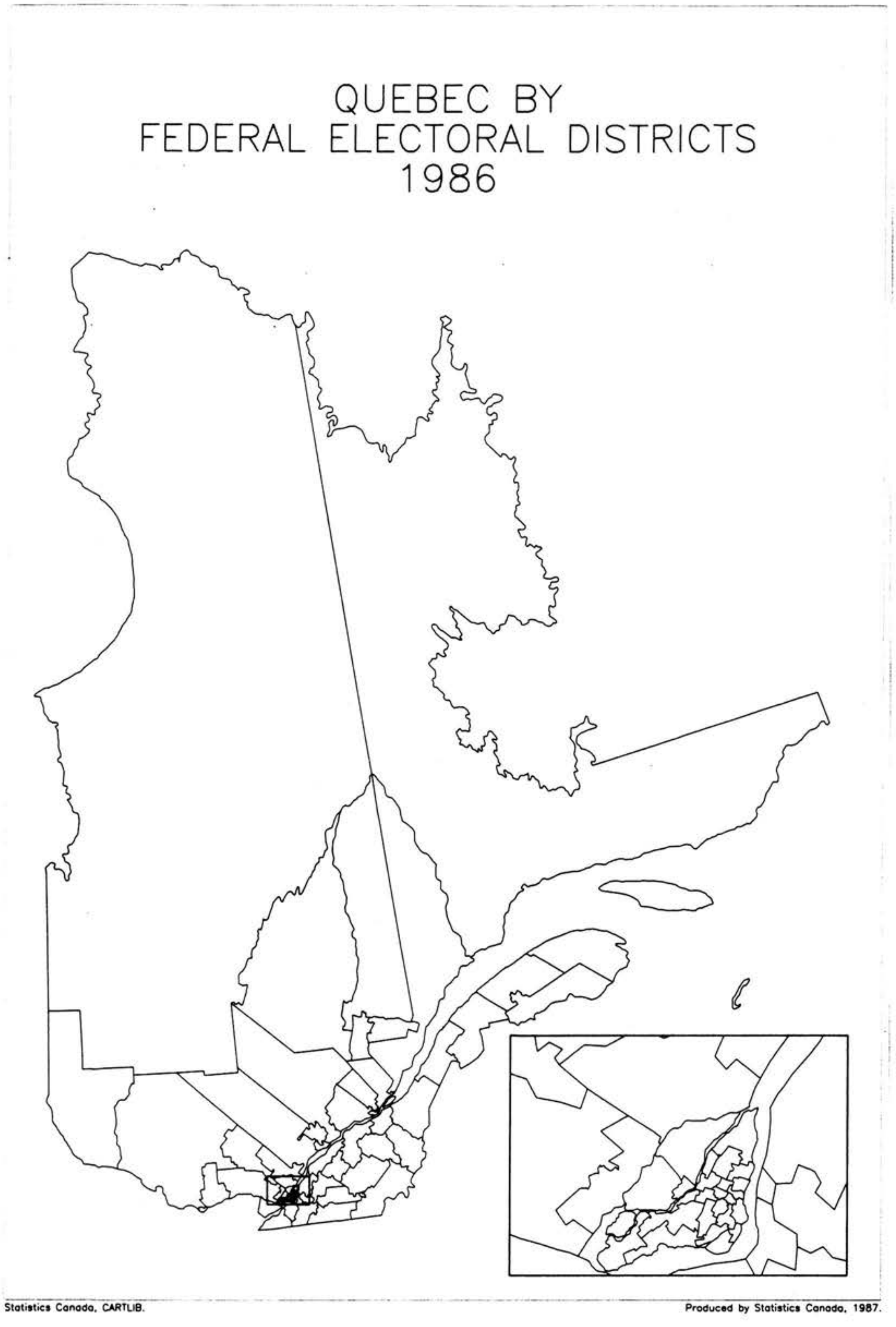


FIGURE 3.5 FEDERAL ELECTORAL DISTRICTS OF QUEBEC



2. **Recognizable Limits**

To ensure that the limits of collection units are well and completely identified, and are clearly visible during field collection operations, they must respect visible physical features such as streets, railway and power transmission lines, or water bodies, such as lakes and rivers, to the highest degree possible. These features must be, therefore, part of the digital cartographic data base if they are to be respected. In urban areas with moderate to high settlement densities, street blocks are respected whenever possible (i.e., avoid splitting street blocks).

3. **Easy Accessibility**

Since the effort required to carry out field collection activities is directly related to the length and ease of travel between census dwellings, 'ease of access' is an important consideration in the districting process. Collection units with poor road access - perhaps due to physical barriers such as bodies of water or exclusive transportation corridors (e.g., freeways, railway yards, etc.) are redistricted to allow for easier access.

4. **Language Homogeneity**

Three types of collection units are recognized:

a. Bilingual Collection Units - where the mother tongue of 10% or more of the dwellings is English and 10% or more of the dwellings is French.

b. Unilingual English Collection Units - where the mother tongue of 10% or more of the dwellings is English and less than 10% of the dwellings of the collection unit is French.

c. Unilingual French Collection Units - where the mother tongue of 10% or more of the dwellings is French and less than 10% of the dwellings of the collection unit is English.

The determination of the linguistic designation of a proposed collection unit by manual methods is complicated whenever collection unit boundaries must be changed from the previous Census. The exception to this statement occurs whenever entire collection units from the previous Census are combined into a single collection unit for the current Census. In this case the percentage of the total dwellings of the new collection unit can be calculated by taking the percentage of a given linguistic group times the same total number of dwellings and multiplying this product by 100 and then dividing by the overall total number of dwellings. The more general case occurs whenever parts of former collection units are combined. To simplify manual calculations, a standard ratio (25%, 50% or 75%) is applied to the statistical information (i.e., 50% French and English, 25% French and 75% English, or **vice versa**).

5. **Homogeneous Field Collection Methodologies**

Given the diversity of conditions across Canada, three different methodologies are usually employed in taking a Census:

a. Drop Off/Mail Back

Census questionnaires are dropped off at dwellings by census representatives and returned to census commissioners via the postal system.

This method is usually used in population centres with 10,000 or more persons, or in other centres specifically recommended by the Regional Offices.

b. Drop Off/Pick Up

Census representatives drop off questionnaires and return to collect them when they have been completed.

This method is employed in most rural areas and in smaller urban centres (under 10,000 persons).

c. Canvasser/Interviewer

Questionnaires are completed at the dwellings by the census representative on the basis of interview responses.

This method is applied in very sparsely settled areas, and remote areas, typically with small populations spread over a vast territory.

It is worth noting that a fourth method employed by the U.S. Bureau of the Census in its largest population centres - 'Mail Out/Mail Back' is not currently used in Canada although research into the feasibility and cost-effectiveness of this approach in the Canadian context is under study. If adopted in Canada, this would likely create a new class of workload targets which would result in a significant change in the size and number of collection units.

To simplify field operations, data for individual collection units must be collected using a single methodology.

6. **Historical Continuity**

In order to facilitate longitudinal research studies at the level of collection units, the previous two Censuses in Canada have required that collection unit boundaries be changed only where:

- o recommended by Regional Operations personnel,
- o necessitated by changes in geostatistical limits,
- o required due to significant growth or decline, or
- o requested for special areas by Provincial Governments.

Another major cause is a change in collection methodology (e.g., for the 1991 Census there is an expansion of the use of the 'drop-off/mail-back' method (described later in this section) which increases the average size of the collection units).

Tables 3.5a and 3.5b are standard statistics that are produced by districting staff at Statistics Canada and indicate that, even with such a policy, the total number of changes remains quite large (i.e., only between 52.4 and 54.5 percent remain unchanged).

TABLE 3.5a COLLECTION UNIT STABILITY FROM 1976 TO 1981

Changes in Collection Units from 1976 to 1981 were:

	Number	% of Total
One 1981 EA = two or more 1976 EAs	288	0.7
Two or more 1981 EAs = one 1976 EA	2,411	5.9
One 1981 EA = one 1976 EA	21,604	52.4
Total number of equivalent sets of 1981 and 1976 EAs	27,440	66.6
Total number of enumeration areas	41,197	100.0

TABLE 3.5b COLLECTION UNIT STABILITY FROM 1981 TO 1986

Changes in Collection Units from 1981 to 1986 were:

	Number	% of Total
One 1986 EA = two or more 1981 EAs	4,182	9.5
Two or more 1986 EAs = one 1981 EA	1,472	3.3
One 1986 EA = one 1981 EA	24,000	54.5
Total number of equivalent sets of 1986 and 1981 EAs	26,895	61.1
Total number of enumeration areas	44,042	100.0

7. Equal Workload

Within the constraints listed above, the primary objective of the districting process is to construct a set of collection units which (when grouped as workloads, if necessary) result in a fair and even distribution of the total workload amongst the thousands of census representatives. For the 1986 Census, over 35,000 census representatives were responsible for enumerating the 44,042 enumeration areas. (That is, the 44,042 EAs were grouped into approximately 35,000 'workloads').

Currently, the main yardstick for measuring the distribution of work under the manual districting methodology is the total number of dwellings assigned to each collection unit. To the degree possible, clerical staff responsible for the districting operation are expected to district compact, equitably-sized enumeration area collection units that respect pre-established target numbers of dwellings. These targets vary with the density of dwellings, the number of farms, the difficulty of enumeration, and the method of enumeration.

The types of collection units are based on density of dwellings **per square mile** as shown in Table 3.6a (1981) or **per square kilometer** as shown in Table 3.6b (1986) and are used to determine the rate of pay for the census representatives. (Note that the table values represent areas that are essentially the same size.)

Table 3.6a COLLECTION UNIT TYPES BASED ON DENSITY (1981)

TYPE	DENSITY CRITERIA	
A	250	dwellings or more **per square mile**
B	125 - 249	dwellings per square mile
C	25 - 125	dwellings per square mile
D	2.5 - 24	dwellings per square mile
E	0.5 - 2.49	dwellings per square mile
F	less than 0.5	dwellings per square mile.

Table 3.6b COLLECTION UNIT TYPES BASED ON DENSITY (1986)

TYPE	DENSITY CRITERIA	
A	more than 96	dwellings per **square kilometre**
B	48 - 96	dwellings per square kilometre
C	9.60 - 48	dwellings per square kilometre
D	0.96 - 9.60	dwellings per square kilometre
E	0.19 - 0.96	dwellings per square kilometre
F	less than 0.19	dwellings per square kilometre.

In 1981 and 1986, the following workload targets were set:

Table 3.7a TARGET WORKLOAD LIMITS (DWELLINGS) - 1981 CENSUS

		TYPE OF COLLECTION UNIT				
METHODOLOGY		A	B	C	D	E
Mail Back	Designated Area	225	-	-	-	-
Mail Back	Urban	375	325	275	-	-
Pick-Up	Urban	300	-	-	-	-
Pick-Up	Rural, Non-farm	300	250	200	175	150
Pick-Up	Rural, Farm	-	-	175	150	125
Apt. Bldg.	Designated Area	225	-	-	-	-
Apt. Bldg.	Other	350	-	-	-	-

Table 3.7b TARGET WORKLOAD LIMITS (DWELLINGS) - 1986 CENSUS

		TYPE OF COLLECTION UNIT				
METHODOLOGY		A	B	C	D	E
Mail Back	Designated Area	225	-	-	-	-
Mail Back	Urban	375*	325*	275*	-	-
Pick-Up	Urban	300*	-	-	-	-
Pick-Up	Rural, Non-farm	300*	300*	300*	200*	150
Pick-Up	Rural, Farm	-	-	175	150	125
Apt. Bldg.	Designated Area	225	-	-	-	-
Apt. Bldg.	Other	350	-	-	-	-

* Collection units were not to be redistricted unless they surpassed the target by more than 25 dwellings.
(Note: "Designated areas" are those which are difficult to enumerate.)

3.6 DETERMINING THE QUALITY OF COLLECTION UNIT DISTRICTINGS

The fundamental issue in researching the feasibility of computer-aided creation of districts for census enumeration is how to assess the comparative quality of alternative districts and coverages. Research into current practices in the Canadian Census failed to uncover any existing techniques for this process.

The objective assessment of alternative digitally generated districtings requires an evaluation function that is related to the information available at the time the districts are produced. Information about recent or forthcoming changes in the distributions of fundamental data (streets, dwellings, linguistic groups) could only be considered if revised or estimated values were made available to the program.

Even with these practical limitations on what can be considered in assessing the quality of a districting, there are a variety of factors of varying importance depending upon the various perspectives on quality.

3.6.1 Perspectives On The Quality Of Coverages

The quality of a given coverage can be assessed from a variety of perspectives:

1. **The Census Manager's Perspective**

   The Census Manager is primarily concerned with the likely impact on the **cost** of taking/organizing the census; (under or over) **coverage**; the **timeliness** of the Census results (e.g., ease of assuring quality and performing tabulations -- especially those tabulations required under legislation); and **service** to the public (e.g., the linguistic preference of those being enumerated).

2. **The Census Takers' Perspective**

   The Census Commissioners and Census Representatives are most concerned about the size and distribution of the **workloads** as this has a direct bearing on the elapsed time of the census and subsequently on the hourly income of the Census Representative.

3. **The Historical Analysts' Perspective**

   Historical Analysts want the space sufficiently disaggregated and the data collection units sufficiently stable to facilitate **longitudinal analysis**. It is useful to note that in areas that the Census data have been geocoded to the level of the blockface, these concerns are less significant because they can be handled by post-Census dissemination capabilities.

4. **The Geographic Analysts' Perspective**

Geographic Analysts want building blocks that **conform** to standard regions and zones that are **uniform** or homogeneous (especially in relation to population characteristics) and are of a reasonable **size and shape** (e.g., compact) to be able to form meaningful regions for analysis for varying geographic frameworks (e.g., postal zones). Again, in areas that the Census data have been geocoded to the level of the blockface, these concerns are less significant.

3.6.2 Quality Characteristics Of Collection Unit Districtings

Given the aforementioned districting criteria and varying perspectives, twenty characteristics of districted coverages for census collection are examined to determine the role they play in affecting the ultimate quality of that coverage. The characteristics are given in priority sequence. The methodology used to determine the relative significance of each of these characteristics is presented at the end of this section.

1. **Completeness**

To be considered a 'census', the collection of data must attempt to encompass the entire universe of entities under study. Thus any proposed coverage that does not partition the entire given space is not considered acceptable. Completeness is tested in two ways: firstly, by ensuring that all of the building block entities are assigned to at least one zone in the coverage; and secondly, by intersecting (using polygon overlay) the zone boundaries to ensure that no gaps exist.

2. **Uniqueness**

Inherent in the definition of a census is the notion that the entities are canvassed only once. Further, the census is taken with respect to a very specific time and the respondents are expected to record their information in relation to their location at that time. It is the responsibility of the Census Representative to ensure that dwellings within a given collection unit are canvassed only once. The responsibility for ensuring that all dwellings are assigned to only one collection unit belongs to the person or system performing the districting. Thus, almost as important as ensuring that there is no **under coverage** (i.e., that the coverage is 'complete') is ensuring that there is no **over coverage**. To be acceptable, a proposed coverage must be demonstrated to be unique. Again two tests are applied: firstly, by ensuring that all of the building block entities are assigned to only one zone in the coverage; and secondly, by intersecting of the collection unit boundaries to ensure that no overlaps exist.

3. **Conformance with the Geostatistical Hierarchy**

To facilitate the timeliness and ease of tabulation and checking, collection unit districts are made to conform to a pre-established hierarchy of zones termed 'geostatistical areas'.

Thus, it is important that the proposed coverage conform exactly to the geostatistical hierarchy that is established for the given census. This is achieved by ensuring that the cartographic reference frame that is used to design the proposed coverage is first partitioned by fundamental elements (or "least common areas") of the geostatistical hierarchy. These fundamental elements can be generated by the use of a "polygon overlay" facility to intersect the lower order zones in the geostatistical hierarchy. (Typically, this involves a Census Tract boundary being subdivided (usually infrequently) by a Census Subdivision, Federal Electoral District and/or Urban Area boundary.

4. **Single Methodology**

To standardize and simplify field operations, it is a requirement that collection methodologies (e.g., 'mail-back', 'pick-up' and 'canvassing') not be mixed within a given collection unit. Indeed, methodologies are typically consistent within a given Census Commissioner District (a group of 18-20 collection units supervised by one Census Commissioner in 1986), and thus this criterion can be considered to be satisfied **a priori** for the test sites used for this model.

5. **Historical Continuity**

Longitudinal analysis of a region through time is greatly facilitated if the zones by which the socio-economic data are collected are stable over the given timeframe. The model makes allowances for the added benefits of historical continuity by having a special evaluation function for the re-use of former collection unit boundaries that is less demanding than the regular evaluation function.

6. **Agricultural Considerations Respected**

Agricultural activity affects the taking of a census in a number of ways:

- o The taking of the census is delayed until after the usual start of the planting season (i.e., it is taken nationally in June rather than April as is done in the U.S.A.).

- o The methodology used in agricultural areas (i.e., 'pick-up') is more labour intensive.
- o The number of questionnaires (census of agriculture and census of population) is doubled.
- o For a given density range of dwellings, agricultural activity dictates that districts contain fewer dwellings (given the extra work associated with coincidently taking the Census of Agriculture).

In previous censuses, there was the additional condition that a maximum of 99 farms be contained in any given collection unit. However, this is no longer a strict constraint. In the highly urbanized Census Tracted areas throughout Canada, the effect of agricultural activity considerations is minimal. Therefore, the impact of this factor on the model is limited to the appropriate specification of districting targets in agricultural areas.

7. **Minimum Number of Zones**

The enumeration and field operations represent over one-third of the total cost of producing a census in Canada. A substantial portion of this amount is related to the salaries of the Census Commissioners and Representatives. Minimizing the total number of enumerators needed to collect the census can increase cost-effectiveness through reduced training and other costs that vary with the number of enumerators (e.g., the number of Census Commissioners). The minimum number of zones results if each representative is assigned a maximum workload for the given settlement density.

8. **Respect Visible Features**

To be able to locate and remain within their prescribed territories, it is important that the census representatives be provided with districts that conform, whenever possible, to features visible 'on the ground'. These features typically include streets, railways, rivers and power lines and are, fortunately, recorded in the digital cartographic files for larger urban centres.

9. **Facilitate Accessibility**

Many dwellings are accessible only from a single direction due, perhaps, to the street pattern, physical barriers (e.g., cliffs, streams, fences, etc.) or building entrances. To minimize the effort needed to enumerate the dwellings, it is essential to facilitate accessibility in designing Collection Units. This concern is particulary acute in the case of remote or 'off-shore' dwellings. Since terrain and accessibility variation are not explicitly represented in the digital cartographic data bases, a combination of factors (e.g., use of previous collection units, respecting the street network, and not splitting blocks), when combined with the use of field checks, mitigate against poor accessibility.

10. **Consider Supervisory Workloads**

Total dwelling counts for a given area are seldom multiples of the target dwelling count per district. Further, dwellings are, and should be, grouped by block. This makes it even more unlikely that the optimal **number** and the optimal **size** of districts can be achieved simultaneously. Currently, this difficulty is reduced somewhat by directives which instruct the staff to take into consideration supervisory workloads. In practical terms this means that small residual districts can be formed in a given area (e.g., in a Census Tract) if a complementary small residual district is available from a neighbouring area (i.e., another, ideally neighbouring, Census Tract) supervised by the same Census Commissioner.

11. **Respect Enumerator Workload Limits**

Workload limits, expressed as dwelling count totals for the different levels of dwelling density, types of collection methodologies and for designated and non-designated areas, is one of the principal factors used in manual districting collection units.

12. **Respect Collectives**

Collectives are typically sub-blockface entities and must be respected in order to be considered as a separate statistical universe. A collection unit may not combine collective and non-collective dwellings.

13. **Respect Blockfaces**

Without a visit to the field, it is usually very difficult to decide where to arbitrarily split a blockface. In most cases, splitting a blockface also increases the difficulty of both the enumeration and the geocoding processes. Therefore, if at all possible, blockfaces are not split during the districting process.

14. **Respect Blocks**

Splitting blocks is also avoided whenever possible. However, it is clear that blocks with a total workload in excess of prescribed limits for a single representative must be split.

When splitting blocks is necessary, it is usual to respect blockface units whenever possible.

15. **Respect Contiguity**

To minimize the travel time between parts (e.g., groups of blocks) of a given collection unit work assignment and to facilitate the analysis of the statistical results, adjacent blocks are combined to form contiguous collection units whenever possible.

16. **Respect Linguistic Groupings**

As mentioned earlier, unilingual (English or French) districts are preferred over bilingual districts on the basis of ease of recruitment. Blocks are "swapped", where possible, to convert bilingual districts to unilingual districts.

The result of this procedure may introduce a systematic bias that may exaggerate the degree of linguistic separation if the collection units are also used for dissemination purposes. On the other hand, such units will likely be slightly more homogeneous in terms of population characteristics and cannot be faulted for concealing the degree of segregation (which is the usual source of concern).

17. **Equitable Route Length Allocation**

Since all parts of the street network containing dwellings must be visited at least once, route lengths should be apportioned equitably and in a manner that minimizes the degree of unavoidable route overlap caused whenever several blocks (i.e., more than 4) must be canvassed.

18. **Equitable Distances to Work Location**

Where possible, the effort required to reach the start of a collection unit 'route' should be taken into consideration in the districting process. Census Representatives that must travel proportionally greater distances to reach districts that are less central can be compensated by being paid for travel time and by having relatively less work to do upon reaching their district. In the manual districting process, this factor receives greater consideration in rural settings.

19. **Maximize Homogeneity**

Homogeneous or 'uniform' districts enhance the validity of statistical inferences. The major difficulty in attempting to generate collection units that are homogeneous stems from two factors:

- o Homogeneous zones, based on data from the previous (5 year old) census, are unlikely to have the identical level of homogeneity for the current census.
- o Different applications require that different criteria be used as the basis of the 'homogeneity factor'. Combining criteria tends to weaken the level of homogeneity that can be attained.

The ability to form dissemination units that are sets of blockface units, grouped according to arbitrary, post-Census criteria reduces the need for homogeneity of collection districts.

20. **Strive for Compact Shapes**

Indices for evaluating the shape of irregular polygons have received substantial treatment in the literature (see J. Kimerling, **et al**, 1973 for a review of several measures). It is generally accepted that compact shapes are preferred as they maximize the area covered for a given perimeter or conversely minimize the perimeter for a given area. In the literature on political redistricting, this characteristic is ranked with population equity and district contiguity as one of the three most important requirements. This is due, in part, to the concern that

> "without some requirement of compactness, the boundaries of a district may twist and wind their way across the map in fantastic fashion in order to absorb pockets of partisan support" (Reock, 1961, p. 71).

Fortunately, this concern about possible **gerrymandering** is not of direct concern to the census collection unit districting problem as the units are relatively small and as the Census Representatives have little say as to the districts to which they are assigned nor to the size and shape of the districts that are generated (unless it can be shown that the district has to be split due to major unexpected growth in settlement densities). (Since political districting often tends to respect former collection unit boundaries -- as this is the unit by which the counts are typically provided -- there might be some concern about downstream impacts on political redistricting if data other than language and dwelling counts were available to the collection unit districting process.)

In addition, the results shown in Table 3.8 of an analysis of Figure 3.6 provide some quantitative insight into why this typically important characteristic of the quality of a proposed districting is ranked so surprisingly low for collection unit districting.

Figure 3.6 ALTERNATIVE COMBINATIONS OF SIMILAR BLOCKS

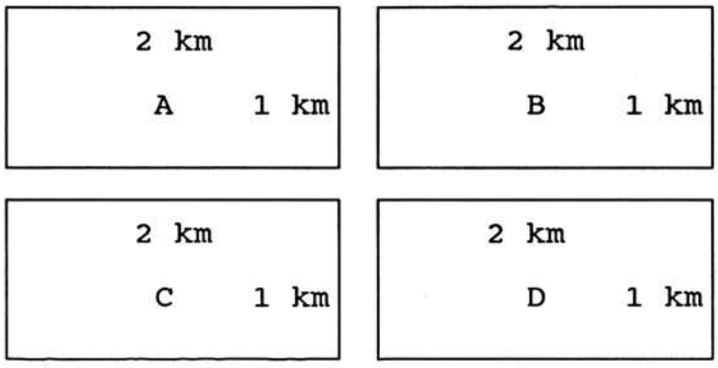


Table 3.8 SHAPE CHARACTERISTICS OF FIGURE 3.6 BLOCK GROUPS

BLOCK GROUP	DIMENSIONS	AREA	EXTERNAL PERIMETER	TOTAL PERIMETER
A	2km x 1km	2kmkm	6km	6km
B	2km x 1km	2kmkm	6km	6km
C	2km x 1km	2kmkm	6km	6km
D	2km x 1km	2kmkm	6km	6km
AB	4km x 1km	4kmkm	10km	12km
CD	4km x 1km	4kmkm	10km	12km
AC	2km x 2km	4kmkm	8km	12km
BD	2km x 2km	4kmkm	8km	12km

It is clear from Table 3.8 that while the block group AC is more compact than AB and therefore the external perimeter is reduced by 2 kms, there is no difference or advantage in the total perimeter that must be traversed to canvass all of the constituent blockfaces. Thus, little explicit emphasis need be placed on shape by the autodistricting capacity.

A summary of varying perspectives on each of the quality characteristics from above is presented in Table 3.9. The assessments were compiled on the basis of discussions with individuals familiar with the process of creating collection unit districts for the Census. The ranking that results from the analysis of these preferences was produced solely as a guide to the implementation of the model and should not be considered as the definitive statement on the relative merit of each quality characteristic. All but the homogeneity characteristic (which can be implemented via geocoding) will be included in the current implementation of the model.

Table 3.9 CHARACTERISTIC PREFERENCE TABLE

Characteristics	Census Manager	Census Takers	Census Data Users	
			Historic	Geographic
Completeness	++	++	++	++
Uniqueness	++	++	++	++
Hierarchy conformant	++	++ *	++	++
Single methodology	++	++	+	+
Respect Agricultural limits	++	+	=	=
Historical Continuity	++	+	++	+
Minimum number of zones	++	++	--	--
Respect visible features	+	++	++	++
Facilitate Accessibility	+	++	=	=
Consider Supervisor Areas	+	++	=	=
Respect Workload Limits	+	++	=	=
Respect Collectives	+	++	=	=
Respect Block Faces	+	+	++	++
Respect Blocks	+	+	++	++
Respect Contiguity	+	+	+	+
Respect Linguistic Groups	=	++	++	++
Minimize route length	=	+	+	+
Minimize start distance	=	+	=	=
Maximize homogeneity	=	=	++	+
Strive for compact shapes	=	=	+	+

LEGEND:
++ STRONGLY SUPPORTS -- STRONGLY OPPOSES
+ supports somewhat - opposes somewhat
= indifferent
* not strictly required for collection alone

Table 3.9 indicates that while there are a few characteristics for which there is general agreement on relative importance, there are much greater number for which there is either disagreement or indifference. The major issue with respect to these 'perspectives' is which are permitted to dominate and to what degree. The methodology used to implement these characteristics is presented in Chapter 4.

3.7 SUMMARY

This chapter has considered at some length the various dimensions (including characteristics, constraints and quality considerations) of the collection unit districting problem in Canada. It briefly documented the spatial frameworks and methods employed in the current manual districting approach. Further details are provided in Appendix C.

The following steps were followed in order to achieve the research objective:

1. Current manual districting procedures were studied through a review of available documentation and hands-on training;

2. The literature was reviewed and the districting staff were interviewed to compare districting objectives, approaches, techniques and evaluation criteria;

3. The staff responsible for early attempts at computer-aided districting at the U. S. Bureau of the Census were interviewed; and

4. A suite of districting tools was developed and subsequently imbedded within an existing geographical information system (GIS), ARC/Info, since none of the existing tools were able to respond to the varying degrees of data availability.

3.8 CONCLUDING REMARKS

The general collection unit districting problem is probably most related to the classical political redistricting problem which is also multi-objective and has a strong orientation towards the equity condition. Other similiarities include:

1. the **completeness** constraint - so that no one is disenfranchized;
2. the **contiguity** constraint - so that the politician represent a single area;

3. the **consistency** condition - so that the districts conform to higher order units such as Provinces or States, and
4. the **similarity** condition - so that the districts are relatively homogeneous to counter any attempts at **gerrymandering**.

This latter concern tends to complicate the political redistricting process since it tends to require an iterative, negotiation-oriented approach be taken.

While Census collection unit districting is not fraught with the 'political' elements of the political redistricting problem, there are several elements which make it, in some senses, a more difficult problem including:

1. the size of the problem (at the national level in Canada, political districting involves the creation of less than 300 districts while over 40,000 Census collection units must be created);
2. the size of the fundamental building blocks for political redistricting tend to be fixed (often they are enumeration districts or areas) while for the census collection unit districting problem they are variable (i.e., sub-blockfaces, blockfaces and blocks);
3. the **consistency** condition applies to both the district level (i.e., to maintain former districts) as well as at the level of the higher order units. There is also a cartographic manifestation of this condition that insists that the districts should respect features that are visible on the ground;
4. the **similarity** condition applies to both language distributions and to collection methodology types;
5. the **efficiency** condition has greater significance in the case of Census collection unit districting since it can be tied directly to out of pocket costs for Census operations; and
6. the **centrality** condition becomes increasingly relevant in rural areas where the dwelling density drops and the size of the districts grows dramatically.

These conceptual differences have a direct impact on the suitability of designs of alternative solution procedures as will be discussed in the next chapter. The over 100 fold impact in cost differentials for generating districtings means that tradeoffs between the cost and the quality of final results tend to be more heavily weighted in the direction of cost concerns than might be the case for political redistricting. There is a comparable tradeoff that must be made in terms of timeliness since collection unit districting must be completed within a fixed timeframe every five years. Political districting in Canada, on the other hand, has a much longer and more flexible timeframe. Here again, the quality component must be subordinated to the timeliness concern in the case of Census collection unit districting.

Thus it can be expected that a solution procedure appropriate for collection unit districting will:

1. be less interactive;
2. have more components or subsystems;
3. satisfy more criteria; and
4. be more sensitive to cost and throughput concerns.

The nature of the implemented solution is described in some detail in the next chapter including:

a. a methodology for employing the existing and enhanced set of GIS capacities;
b. a set of objective functions to measure the relative quality of individual collection unit districts;
c. dispersion indices to guide the selection of alternative districting tools; and
d. an evaluation function to assess alternative districtings for the case study areas.

The empirical testing conducted on each districting method is then described in Chapter 5.

CHAPTER 4

MODELLING STRATEGY, METHODS AND OPTIONS

4.1 PURPOSE OF THE CHAPTER

The purpose of this chapter is to describe in some detail the strategy that has been employed for the design and development of a multi-stage districting model together with the constituent components of that model (i.e., the modules, approaches, steps and methods that have been implemented as a prototype districting system). Both the individual components of the model and their interrelationships to one another are described.

4.2 STRUCTURE OF THE CHAPTER

The chapter begins with a discussion of the modelling strategy including the objectives, assumptions and design characteristics. This is followed by an overview of the entire theoretical model. The main components of the theoretical model and the implemented system are first introduced in the context of variable data availability together with their interrelationships (depicted in Figure 4.2). This is followed by an accompanying overview of the usual operation (or "process flow") of the model/system and a description of each of the elements of the model in greater detail. Finally, a summary of the decision tables/trees and options is provided.

4.3 MODELLING STRATEGY

The strategy selected for building an appropriate and comprehensible model was based on a number of objectives and assumptions.

The main objectives were that the districting model/system (including, where appropriate, the selection of districting methods and the evaluation of results) be:

1. highly automated, since it is clear from the literature [e.g., Thalmann, et al, 1982] that the straightforward emulation of the manual districting processes and results by simply

using manually driven interactive techniques unreasonably constrains the approaches taken and probably is too limited an objective for research at the doctoral level;

2. cost-effective, which is related to a considerable degree to the previous point, since it is evident from early work by the U.S. Bureau of the Census, [Bonnette, 1980], that processes making extensive use of interactive revisions may result in operational systems not being applied in a production environment;

3. able to cope with highly varying amounts of digital information so that its use can eventually be extended beyond the large urban centres for which considerable digital information is currently available;

4. able to consider all of the districting criteria that were discussed in Chapters 2 and 3; and

5. able to objectively assess and report on the quality of: (a) the previous manually generated collection unit districts; (b) the districts generated by the various districting tools; and (c) the districts that have been generated by the various districting tools and then enhanced by manual intervention.

The assumptions underlying the selected strategy include that:

1. replicating, to the highest possible degree, the elements of the manual districting process in a spatial decision support system that goes beyond providing a straight forward interactive computerized environment, would not only lead to insights that could provide a foundation for an eventual implementation of an expert system for autodistricting, but would also ease the assimilation of the technology into the production environment since most of the elements of the model would be familiar or seem intuitively sensible to the end users;

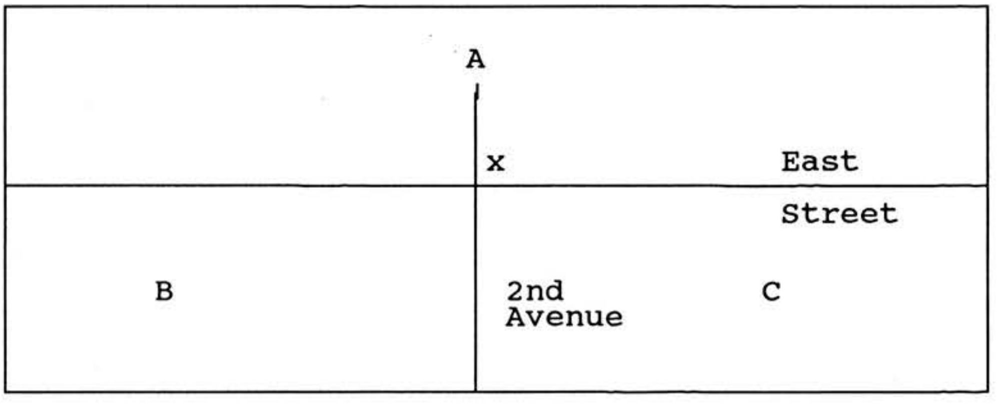
2. because the general timeliness of the data traditionally used to generate preliminary collection unit districts is poor (typically five years old), and because the results of the districting process (manual or automated) undergo a field check prior to the Census to compensate for the untimeliness of the input data, operationally adequate districts are likely to be far less than optimal. (The attainment of optimal solutions is the goal of most of the computationally intensive, i.e., more expensive, districting systems and is beyond the minimum requirements of the current districting process. See Eilon, 1972, for a comparison of optimizing and satisficing strategies);

3. a set or "tool kit" of simple districting methods would be able to handle the variability of settlement densities, patterns of visible features (on the ground) and accessibility considerations (e.g., highways that obstruct what might appear to be through streets);

4. limitations of time and resources precluded the design, implementation and testing of a large number of districting methods for the tool kit or spatial decision support system;

5. adopting a route-oriented districting model [Scott, 1971] (that might generate near-optimal allocations) would not provide the necessary sense of territory essential to completing the enumeration of the country (i.e., the current orientation usually results in Census Representatives being responsible for the enumeration of all of the dwellings within a standard "city block" and thereby reduces the likelihood of overcoverage or undercoverage). For example, in an area based approach, all of the dwellings in the "city block" labelled 'A', in Figure 4.1 below, including the cul-de-sac at the top end of 2nd Avenue, are the responsibility of the Census Representative enumerating that "city block". A network approach might generate a result, however, that could lead to confusion between the responsibilities of that Census Representative and the responsibilities of the Census Representative(s) enumerating the two sides of the remainder of 2nd Avenue if the cul-de-sac was assigned as part of 2nd Avenue rather than as a side of "city block" 'A'. An area based approach also reduces confusion about responsibility for a dwelling that might be located at a point 'x' with entrances on both 2nd Avenue and East Street.

Figure 4.1 Area Versus Route-Oriented Districting Example



6. attempting to combine "optimal" district coverages with optimal district routings would overly complicate the districting process and greatly increase the computing resources

required; and

7. computationally more elaborate techniques such as linear programming [Marlin, 1981], following space-filling curves [Mandelbrot, 1977 and 1983] or random [Deichsel and Tubingen, 1980], or shortest [Pollack, et al, 1960] paths, automated "swapping" of components of the previous collection unit districts, and/or incorporating Theissen polygons [Rhynsburger, 1973] or clustering algorithms [Openshaw, et al, 1982] might, eventually, be added to the repertoire of districting tools (when the cost of computing drops further).

In light of these strategic assumptions, the greatest promise for successful collection unit districting that is flexible and cost effective at the national level (i.e., for the entire country) appears to be a **multi-stage, multi-component model** that combines a set of **disaggregation, aggregation and re-configuration approaches**, while allowing for the **re-use** of previous collection unit districts.

The **design** of such a multi-stage model for districting collection units for the Canadian Census has involved:

1. the **conceptualization** of a theoretical framework capable of being applied (eventually) to all of Canada, in spite of the current variable availability of digital data;

2. the **design** of a process for identifying and extracting (or "filtering") natural units (of sufficient size to be considered individual collection units) at various levels of resolution (or geographic scale);

3. the **development, selection and enhancement** of a set of easily understood districting techniques (or "tools in a tool kit") that can be applied (by non specialists in the field of electronic data processing, EDP) to varying types or configurations of **geographic building blocks**; and

4. the **incorporation** of human intervention by districting specialists, when appropriate and cost effective.

The **strategy** that has been followed for **implementing** the multi-stage, multi-component model outlined above appropriately recognizes the **practical limitations** of time and cost related to doctoral research.

Consequently, while the thesis attempts to make contributions on both the **theoretical** and the practical level, due to limitations of time and cost, it is understandable that it has not been possible to put all of the theoretical ideas into practice. Nevertheless, they are outlined and discussed, where appropriate, because of the conceptual insights they yield.

Thus, the approach taken has been to implement and test the most useful and essential components first and relegate the remaining components to post-doctoral work, though these components are often straightforward to implement. To provide a clear picture of achievements to date, however, theoretical components which are not yet fully implemented are flagged (parenthetically).

Though not exhaustive, this practical approach has proven more than adequate for the current application of the model.

4.4 DATA AVAILABILITY

As indicated in the preceding chapter, the availability of digital data is the major determinant of how the individual modules of the model/system are employed. The geographic data include the boundaries of the higher level geographical units that, when overlaid, form the boundaries of the least common geographic areas called "Census Tract Parts" or CTPs (which must always be provided) and/or the boundaries of the previous collection units (which may or may not be available). When available, the cartographic data include the streets, railroads, rivers, transmission lines, etc., that, when overlaid, form the boundaries of the least common polygons or "**blocks**". (Bolding is used to differentiate these "**blocks**", from the standard notion of a "city block", which is bounded on all sides solely by components of the street network. The **bolding** of the term **perimetre** is used to indicate that the value refers only to the length of the **block** sides that are streets.)

The statistical data includes the dwelling counts at the collection unit level from the previous Census (which are always available) and the dwelling counts at the level of the blockface (which, when available, are also "rolled up" to the level of the **block**) or the sub-blockface in the case of collectives and special collection units (which also may or may not be available). Finally, statistical data may include actual and/or estimated work achievement rates associated with the components of effort needed to collect the Census questionnaires. When available, these rates can be combined with characteristic values (e.g., the **perimetre** or area of **blocks**) derived from elements of the street network or supplied as input together with the **block** centroids.

While the availability of the individual types of data in digital format is highly variable, there are basically four main CASES. They are considered below, in ascending order of information content:

CASE #1: Only the Census Tract Part boundaries are available;

CASE #2: The previous collection unit boundaries, and the block centroids within the CTP are also available;

CASE #3: The perimetre and/or areal extent of blocks are also available or can be derived from the supplied cartographic network; and

CASE #4: The elapsed time and/or dwelling count values for each block are also available (or can be calculated from the work achievement rates and the other types of data available in CASE #3).

Not only does the use of the modules of the districting model vary for each of these CASES, so too does the form of the objective functions and the procedures for selecting districting methods.

4.5 OVERVIEW OF THE DISTRICTING MODEL

The districting model consists of four MODULES:

1. FILTERING;
2. SELECTION;
3. BLOCK GROUPING; and
4. BLOCK or SUBDISTRICT RE-GROUPING.

Subdistricts are groups of blocks that are contiguous. A district may have one or more subdistricts as a result of a centroid-based decomposition or composition process. However, districts formed by processes that are based on topology ensure that contiguity constraints are adhered to and do not, therefore, generate districts with discontiguous subdistricts.

Each MODULE, and its relationship to the availability of data, is now discussed briefly.

o FILTERING

In either of the last two CASES (i.e., #3 and #4), it is possible to identify units which are already large enough to meet the districting criteria (i.e., in terms of perimetre lengths of streets, areas, dwelling counts or elapsed time values). To simplify the subsequent districting processes, it is expedient to filter these natural districts from the data base as part of the first pass of the model. These natural districts can be filtered out at each of the three nested levels described in Chapter 3, i.e., sub-blockfaces, blockfaces or blocks (hence the use of the term "cascading") if they meet the minimum threshold for a collection unit.

No attempt is made at this stage to combine other units at the same level to form districts.

1. Sub-blockface level
   Collectives and other sub-blockface collection units such as large apartment buildings are traditionally maintained as collection units and, following the 1986 Census convention, are filtered from the data base regardless of size. The affected blockfaces are each flagged with the number of constituent sub-blockface collection units.

2. Blockface level
   Blockfaces that meet or exceed the minimum threshold for collection units are filtered from the data base leaving a remainder value (possibly equal to zero) which is assigned to a duplicate ('residual') blockface (i.e., a blockface with the same centroid but with a dwelling count equal to the remainder value).

   Blockfaces that are too small are left for subsequent "roll-up" to the block level together with 'residual' blockfaces containing remainder values from the blockface FILTERING process.

3. Block level
   Blocks that meet or exceed the minimum threshold for collection units are filtered from the data base leaving a remainder value (again, possibly equal to zero) which is assigned to a duplicate ('residual') block (to preserve the topology of the network). Blocks that are too small (including zero and non-zero residual blocks) are left for subsequent SELECTION or BLOCK GROUPING.

o SELECTION

SELECTION seeks to partition the set **blocks** within the CTP into districts based on the relationship of the **block** centroids to either the **boundaries** of the previous collection units or to the **geometries** of the grid or seed templates described in Chapter 2. In both cases it essentially involves a "top-down" decomposition or DISAGGREGATION process.

(a) The **first** SELECTION APPROACH partitions the set of **block** centroids by their relationship to the boundaries of the previous collection units (which is determined by a point-in-polygon procedure). If the collection units so formed are not too large, relative to the **workload target values**, they are reused. Otherwise, the CTP is redistricted. The prototype system follows the 1986 Census convention that previous collection units with very small workloads are maintained to allow for historical comparability and to permit actual 1986 Census districting results and autodistricting results to be compared. (After deliberations, it was decided that the production system for the 1991 Census would reject the **previous** collection units if they did not meet the same criteria as are applied to the **new** districts -- i.e., if they are too small they will not be reused.) A minimal requirement for the use of the first approach is, therefore, the availability of the **block** centroids and the previous collection unit boundaries in digital format (i.e., minimally data CASE #2).

(b) The **second** SELECTION APPROACH seeks to define new collection units by partitioning **block** units. This is generally based on the location of the **block** centroids relative either to the cells of fixed grids or to seeds which are **fixed** or **random**. The exception is for data availability CASE #1 where the grid templates are used to partition the CTPs without reference to centroids using standard polygon overlay techniques. These "top-down" **DISAGGREGATION models** are based on the geometries of rectangular or circular **grids** or on **seed** locations which, though fixed on a relative basis, have been adapted to the shape of the Bounding Box.

Because the SELECTION of blocks is based solely on the location of a single representative point, the block centroid, the result of this provisional SELECTION can be discontiguous, oversized or undersized (relative to block count, perimetre, area, dwelling count or elapsed time values). Up to three of the nine DISAGGREGATION methods are used on any given Census Tract Part. (While the decision to select at most three tries is arbitrary, in theory it seemed that the square root of the number of alternatives represented an upper bound which grows slowly with significant increases in the number of methods and - as will be evident from the results presented in Chapter

5 - has proven in practice to be cost effective and to have a high probability of identifying one of the best three methods.)

The choice of method depends on the availability and the distributions of requisite digital data. The mechanism for choosing between DISAGGREGATION methods for the various data availability cases is described later in this chapter (together with specific districting evaluation procedures for each data availability CASE).

DISAGGREGATION works best when the distribution of units (and their associated characteristics) is uniformly spread within the CTP or when that distribution is spread in a locally balanced, but perhaps non-uniform, fashion. As the amount of information about the distributions within the Census Tract Parts increases, AGGREGATION is preferred over DISAGGREGATION because it is better able to respond to a variable distribution. AGGREGATION builds the districts "bottom up" from those distributions and distorts the grid and seed templates as required to 'capture' the desired workload. Consequently, the current solution procedure does not perform DISAGGREGATION SELECTION if CASE #4 data are available. Instead, BLOCK GROUPING or "bottom up" AGGREGATION is employed.

o BLOCK GROUPING

BLOCK GROUPING is applied whenever block centroids are available (i.e., CASES #2, #3, and #4). It is the process of creating districts by distorting the shape of the same grid and seed based templates used for DISAGGREGATION so that the sizes (in terms of the block counts perimetre, area, dwelling counts, or elapsed time values) of the districts (with the exception of unavoidable "residual" districts) fall within the target range. This is in contrast to the SELECTION APPROACHES where the pre-defined model templates do not conform to the actual distribution of the workload elements and hence the sizes of the districts (in terms of the characteristic values), can vary dramatically and may bear little resemblance to the target range.

The nine templates or TYPES of districting METHODS for both DISAGGREGATION and AGGREGATION are referred to as ASSIGN processes because the workload elements (or individual blocks are allocated (or 'assigned') on the basis of their relationship to the fixed or variable templates rather than on their topological relationships with one another.

Up to three of the nine AGGREGATION methods are used on any given CTP in accordance with a method selection procedure that varies with the data availability CASE. This is the same general approach as was followed for the DISAGGREGATION methods.

The provisional districts formed by this grouping of blocks are compared with acceptance criteria and the process is complete if the results fall within the permitted thresholds. If BLOCK GROUPING is not successful after three tries, BLOCK or SUBDISTRICT RE-GROUPING is performed.

o BLOCK OR SUBDISTRICT RE-GROUPING

RE-GROUPING is performed, as required, at the level of blocks (for RE-ASSIGNs and ADJUSTs) or at the level of subdistricts (for ANNEX processes) in three STEPS:

1. the ANNEX process is applied at the level of subdistricts emanating from the ASSIGN process to form more contiguous districts than are generated by the unsuccessful ASSIGN process (which uses only block centroids and the associated characteristic values and does not consider the topology of the blocks in forming districts and hence tends to occasionally generate districts with one or more discontiguous subdistricts);

2. the RE-ASSIGN process is applied at the level of individual blocks and moves iteratively outward from a set of cores (whose selection is described below) and allocates blocks to contiguous districts using the topology of the blocks together with the target value/range (though during the testing of the prototype model/system this value/range was exceeded if necessary to maintain contiguity); and/or

3. the ADJUST process which either permits 'exemptions' from the standard criteria or changes the allocation of individual blocks (via block swapping or splitting), through manual intervention, to other districts to overcome pathological data configurations or minor shortcomings in the results from the ASSIGN, ANNEX or RE-ASSIGN processes.

Each of these STEPS is now described briefly and in greater detail later in this chapter (Section 4.7).

The ANNEX Process

Discontiguous parts of provisional collection unit districtings, termed 'subdistricts', formed by DISAGGREGATION or AGGREGATION ASSIGN processes are 'annexed' by larger, adjacent collection units (districts or larger subdistricts), provided the total workload would not become prohibitive. The result is a set of more contiguous collection units that are usually fewer in total number. Details on how the ANNEX process identifies which subdistricts are adjacent and which subdistricts are 'annexed' to one another is left until later in this chapter. It is interesting to note, however, that the use of the ANNEX module frequently generates the highly desirable result of an acceptable districting with fewer than the target number of districts.

(The current implementation of the ANNEX module is only invoked if dwelling counts are available, i.e., data availability CASE #4. Later versions will permit 'annexing' as long as the cartographic representation of the blocks is available, i.e., CASE #3. The author wishes to acknowledge the cost-effective suggestion of a member of the CADP implementation team that, for the 1991 Census production system, the ANNEX process also be applied to the districts from the previous Census.)

The RE-ASSIGN Process

The purpose of the RE-ASSIGN process is to completely overcome the potential for discontiguities that can be generated by the ASSIGN process. To do so, the RE-ASSIGN process begins by establishing the adjacency relationships or "topology" for all of the blocks for any CTP that was not successfully districted by a combination of ASSIGN and ANNEX processes.

The RE-ASSIGN process then takes advantage of the information learned about the distribution of blocks and their associated characteristic values (i.e., dwelling counts or elapsed time values) during the ASSIGN and ANNEX processes by selecting the desired number of core blocks (equivalent to the target number of districts) from the central blocks of the least unacceptable districting from the ASSIGN and ANNEX processes. Specifically, the core blocks are selected by taking the centres of gravity of the block centroids for each of the districts in the least unacceptable districting result from the ASSIGN and ANNEX processes and identifying the nearest block (by proximity to the block centroids of blocks that have not as yet been identified as cores to account for the case where more than one centroid falls within a given block).

The blocks that are adjacent to these cores are iteratively assigned to form contiguous districts. (Since the number of districts is fixed and since the districts are necessarily contiguous, the size of individual districts may exceed the target value/range in pathological cases.) The details about how these iterative assignments are made are provided in Section 4.7.

If the result from the first RE-ASSIGN process is not acceptable, it is also applied to the next least unacceptable result from the ASSIGN and ANNEX districting processes.

The ADJUST Process

Unacceptable collection unit districtings from the ASSIGN, ANNEX or RE-ASSIGN processes are either granted an 'exemption' from meeting one or more elements of the districting criteria or are 'adjusted' to meet all criteria by districting specialists who typically intervene by directing the swapping of blocks between districts (since block splitting, though technically feasible, is discouraged). A detailed description of the ADJUST process is also provided later in this chapter. It is important to note, however, the intuitive nature of all of these heuristic procedures makes it difficult and/or lengthy to provide complete rationales for each element of the solution procedure. Heuristic approaches remain as much an art as a science and are based on a pragmatic underlying philosophy related to tradeoffs with respect to the quality, throughput and cost of the results. They are ultimately justified, or not, on the basis of performance.

Scott [1971, page 39] has pointed out that:

> Heuristic programming is less a rigidly defined mathematical procedure than a very general problem-solving philosophy. Heuristic programming algorithms represent sets of rules which produce solutions to given problems, but which do not necessarily produce the best possible solutions. These rules may be as rigid or as flexible as seems appropriate to the given problem ... The notion of heuristic programming embraces a wide and heterogeneous variety of computational processes ranging from simple trial and error on one hand to elaborate computer search procedures on the other hand.

Scott [1971, page 56] goes on to state that:

> It is to be stressed that it is quite impossible to make any generally valid statement as to the standard of performance of known heuristic algorithms of any type in the solution of very large and very complex problems ... work by Heller (1960) and Scott (1968) would seem to indicate that mere random sampling will, with high probability, produce solutions to given combinatorial problems within ten to twenty percent of optimality. Almost any well-constructed heuristic algorithm will certainly produce better results than random sampling alone ... it may, in addition, be observed that trial and error guided by human judgement and guesswork can sometimes produce good results in the solution of combinatorial problems. This is perhaps particularly so in the case of those problems which exist in geographic space. No doubt this characteristic may result in large part from the fact that such problems can always be expressed cartographically and solved by graphic manipulation.

The STEPS in this solution procedure involve trial and error, graphic manipulation, common sense and some guesswork. These are also elements of the traditional manual districting process. It is a problem solving philosophy that, as Scott [1971, pages 164

- 165] concludes:

> respond{s} well to the sorts of ambiguities which are so strongly characteristic of actual planning situation. Moreover, the philosophy of heuristic programming, making, as it does, practical compromises between computational costs, arithmetic tractability, and the need to achieve optimality, is in a sense a working model of the kinds of compromises and practical adjustments which must be made in day to day planning operations.

Given this context, the remainder of this chapter attempts to keep a reasonable balance between the amount of description of the components and elements and of the rationale for their selection. The description of the rationales is, therefore, carried to the point at which the selection gives intuitive or common sense.

4.6 OVERVIEW OF THE NORMAL OPERATION OF THE MODEL

The previous section described the various processes which might be thought of as "leaves" of a tree representation of the model/methodology.

The following section describes the "set of branches" of that tree that are traversed for a typical application. Later sections will focus on the details of the criteria used to select between alternative paths. A less graphic metaphor would be to say that the flow of description is to define:

1. the set of process sets;
2. the groups of process sets that are typically invoked; and
3. the operators that generate alternative sequences or groupings of process sets.

Figure 4.2 encapsulates the main elements described earlier together with the branching conditions which determine which parts of the module will be executed in each data CASE. The initial application of the districting model for the 1991 Census of Canada is in areas where all of the requisite data are available (i.e., CASE #4). The descriptions which follow Figure 4.2 describe the flow that takes place for the initial (i.e., CASE #4) application of the model.

Unless otherwise indicated by branching statements, the flow of the districting procedure is line by line from the top of the page to the bottom.

FIGURE 4.2 OVERVIEW OF THE COLLECTION UNIT DISTRICTING PROCEDURE

```
CASE 1: Census Tract Part Boundaries Only Are Available              GO TO [C]
CASE 2: Block Centroids and Previous Boundaries are Also Available   GO TO [B]
CASE 3: Perimeter and/or Area Values are Also Available              GO TO [A]
CASE 4: Dwelling and/or Time Values are Also Available
[A]  MODULE: "FILTERING"
       LEVEL: Sub-blockface Units
       LEVEL: Blockface Units
       LEVEL: Block Units
         IF FINISHED GO TO [F]
[B]  MODULE: SELECTION
       APPROACH: Re-Use Previous Collection Units
         IF FINISHED GO TO [F] or IF CASE #4 GO TO [D]
[C]    APPROACH: Disaggregation
         STEP: ASSIGN (METHODS determined by CASE)
           METHOD: GRID
             FORM: RECTANGULAR
                 TYPE: Unidirectional (M9)
                 TYPE: Bidirectional (M6)
             FORM: CIRCULAR
               SUB-FORM: UNIDIRECTIONAL
                 TYPE: Sectors (M2)
                 TYPE: Rings (M3)
               SUB-FORM: BIDIRECTIONAL
                 TYPE: Sectors and Rings (M4)
                 TYPE: Rings and Sectors (M5)
         IF CASE #1 GO TO [F]
           METHOD: SEED
                 TYPE: Extrema-based (M1)
                 TYPE: Regular (M7)
                 TYPE: Random (M8)
         IF FINISHED GO TO [F]
[D]  MODULE: BLOCK GROUPING
       APPROACH: Aggregation
         STEP: ASSIGN (METHOD determined by CASE)
           METHOD: GRID
             FORM: RECTANGULAR
                 TYPE: Unidirectional (M9)
                 TYPE: Bidirectional (M6)
             FORM: CIRCULAR
               SUB-FORM: UNIDIRECTIONAL
                 TYPE: Sectors (M2)
                 TYPE: Rings (M3)
               SUB-FORM: BIDIRECTIONAL
                 TYPE: Sectors and Rings (M4)
                 TYPE: Rings and Sectors (M5)
           METHOD: SEED
                 TYPE: Extrema-based (M1)
                 TYPE: Regular (M7)
                 TYPE: Random (M8)
     MODULE: BLOCK OR SUBDISTRICT RE-GROUPING
         IF CASE IS LESS THAN 3 GO TO [E]
         STEP: ANNEX -- IF FINISHED GO TO [F]
         STEP: RE-ASSIGN -- IF FINISHED GO TO [F]
[E]      STEP: ADJUST
[F]  FINISH
```

If CASE #4 data are available, the model typically consists of six steps:

1. The model first performs the three stage cascading process of **FILTERING** outlined above and described in greater detail later in this chapter, and thereby essentially reduces the districting problem to **selecting, grouping or re-grouping blocks** that are individually too small to be complete collection units;

2. The boundaries of the previous collection units are used to decompose or **SELECT** the **blocks** into districts within the Census Tract Part;

3. If the result of **SELECTION** by previous collection units is not acceptable, composition, or "**GROUPING**", is undertaken using the **ASSIGN** processes of the AGGREGATION approach. This is followed immediately by an **ANNEX** to reduce the number of districts of that assignment to a minimum. (As mentioned, ASSIGN processes for SELECTION using the DISAGGREGATION approach are not expected to outperform ASSIGN processes for BLOCK GROUPING using the AGGREGATION approach for CASE #4 data and are, therefore, not used for the second and third model testing stage);

4. If the result is unacceptable, a **second** (and if necessary a **third** and final) AGGREGATION ASSIGN (using a different method each time) and ANNEX is performed. (The selection of individual methods and the resulting iteration of the flow is not depicted in Figure 4.2 in the interest of simplicity. These topics are discussed further in the next section);

5. If the result is still unacceptable, a RE-ASSIGN is performed on the least unacceptable (and if necessary on the next least unacceptable) of the ASSIGN and ANNEX processes from the previous steps; and

6. If the result is still unacceptable, an ADJUST is performed to either allow an **'exemption'** for one of the earlier attempts or to **'adjust'** the boundaries by (if necessary splitting blocks and then) swapping blocks between collection units in order to meet all the criteria.

A more detailed description of each step, that includes the rationale for establishing the various conventions and evaluation criteria, now follows.

4.7 THE STEPS OF THE MODEL IN DETAIL

The methods and options of the solution procedure are chosen according to the availability of required data. The FILTERING, previous collection unit SELECTION and RE-ASSIGN steps require a combination of geographic, cartographic and statistical data. In theory, the DISAGGREGATION and AGGREGATION ASSIGN processes and the ANNEX process can operate, if required, on subsets of that data (although the ANNEX process, as currently implemented, also requires all of the data). Thus, data availability is one of the principal criteria employed in choosing between branches of the districting model.

A. FILTERING

Cascading FILTERING processes are performed at the sub-blockface, blockface and block levels whenever the data needed to determine whether these individual units are large enough to be individual collection units (e.g., perimetres, areas, dwelling counts or elapsed time values) are available from the previous census, field checks or administrative records. If such counts are not available (e.g., CASE #1 and CASE #2), the FILTERING step is skipped (since no information is available below the level of the block, and the block information consists only of the location of the centroid).

The three stage FILTERING process is used to identify collection units associated with collectives or other establishments, single blockfaces, or single blocks:

1. Sub-blockface Level

   Criteria for identifying collectives or other special establishments were given in Table 3.3 in Chapter 3. During the 1986 Census these 'special collection units' were kept intact from the previous census regardless of size (though they are manually 'split' if too large). Information about collectives from the previous Census is always available, as are counts for other collection units (e.g., very large apartment buildings) at the sub-blockface level, and this information is used to do the FILTERING.

   Therefore, the first stage in the FILTERING process is to identify such sub-blockface 'special collection units' from the census data base or other sources and maintain their status as collection units.

   If appropriate, the counts of dwellings, etc., are reduced for individual blockfaces by the amount associated with the collectives or sub-blockface collection units.

The sub-blockface dwelling counts are compared with the target value (or range) for the proposed districting. There are two cases:

a) If the count is less than or equal to the target value/range, a sub-blockface level collection unit is formed to preserve the previous collection unit and the blockface count is reduced by the size of the sub-blockface collection unit(s).

b) If the count exceeds the target value/range, one or more sub-blockface level collection units are formed and the count is then reduced by the specified target value (e.g., 350 dwellings) for each collection unit formed.

2. Blockface Level

The second FILTERING stage is to examine each blockface unit to see if it is sufficiently large to be considered a collection unit by itself.

The blockface dwelling counts are compared with the target value/range for the proposed districting. There are three cases:

a) If the count is less than the target value/range, a blockface level collection unit is not formed and the count is left unchanged for subsequent "roll-up" to the block level.

b) If the count falls within the target value/range, a collection unit is formed at the blockface level and the count is set to zero for "roll-up" to the block level. (A check has been added to the production system to see if the block total also falls within the target value/range and if that is the case the block level collection unit takes precedence over the blockface level collection unit.

c) If the count exceeds the target value/range, one or more blockface level collection units are formed and the count is reduced by the target value (e.g., 375 dwellings) for each collection unit formed. The remainder is subsequently "rolled-up" to the block level.

Each blockface is flagged with the number of sub-blockface or blockface collection units that have been formed by FILTERING. Those that were large enough to be collection units are saved for subsequent interactive review and edit (to be able to split the blockface and explicitly locate the boundaries of the blockface collection unit on a map).

Residual blockfaces (perhaps with zero dwelling counts) are preserved to maintain neighbourhood relationships (i.e., the topology) of the network. As noted in the initial overview, these neighbourhood relationships are used by the RE-ASSIGN and the ANNEX processes. A "roll-up" of blockface counts and residual blockface counts to the block level is then performed to generate the block level counts needed for the next stage.

3. **Block Level**

In the third and final FILTERING stage, the blocks are examined to determine if they are sufficiently large to be considered collection units. Block counts that were 'rolled-up' from blockface counts (or remainders) are compared with the target value/range for the proposed districting. Collection units are then formed in a manner identical to that described above for blockfaces.

Blocks that exceed the upper bound of the target range are saved for subsequent interactive review and edit while blocks that are too small pass on to the next step for grouping into districts.

Individual blocks that meet the criterion (i.e., the counts fall in the target value/range) for a district are considered "preliminary districts". These collection units will eventually become final, once any embedded "collectives" or blockface collection units that have been identified at earlier stages have been made explicit. Once the splits are introduced, the finalized collection units are packaged and sent to the field, first for verification as part of the field check process and then, if no changes are required, for the taking of the Census.

'Residual' block units result if the target value is exceeded. For example, for the dwelling count option, a block of 500 dwellings would generate one collection unit of 375 dwellings and leave a 'residual' block unit of 125 dwellings.

The splitting of blocks must, at present, be handled by interactive means. For the moment, each block is flagged with the number of block level collection units that were formed by FILTERING and the remainder is used as the block count for the subsequent stages of the solution procedure.

(A member of the team implementing the 1991 Census production system suggested that block splits from the districts for the previous Census be maintained and, if necessary, used to form part of the boundary of the new districts. While this enhancement does not directly alter the results of the specified FILTERING process, it does reduce the need to introduce splits interactively later in the process and thereby reduces operating costs.)

B. SELECTION AND BLOCK GROUPING

The first SELECTION step is always to check the suitability of re-using the previous collection unit boundaries (i.e., Enumeration Areas or "EAs" from the preceding Census). The criteria for retention of the previous collection unit boundaries varies somewhat with the availability of complementary cartographic (i.e., street pattern) and/or statistical (i.e., dwelling count and elapsed time) data. (The availability of data also affects how the SELECTION and BLOCK GROUPING ASSIGNS are applied.) The criteria and approaches utilized are now discussed for each of the four cases of data availability in ascending order of information available. In all cases, the number and size (in terms of the number of dwellings) of the collection units from the previous Census are also known (as this is a fundamental product of a Census).

1. CASE # 1 - ONLY GEOGRAPHICAL BOUNDARY DATA ARE AVAILABLE.

The boundaries of the fundamental geostatistical units (i.e., Provincial/Census Tracts (PCTs and CTs), Census Subdivisions (CSDs), Federal Electoral Districts (FEDs) and Urban Areas (UAs)) are intersected or 'overlaid' with one another to form least common geographic areas termed 'Census Tract Parts' (CTPs). In this CASE, the target number of districts, T, must be provided to the model (and is usually based on a combination of the results of the previous Census and local area knowledge). The target value, V, is calculated by dividing the surface area, A, of the CTP by the target number of districts. That is, $V = A / T$, which is the target value (in this case area) for each district in square kilometres. (Upper case letters will always refer to values at the level of the CTP and lower case letters will always refer to values at the level of the block or groups of blocks including subdistricts and districts.)

a) Check Previous Collection Units

According to the 1986 Census convention, if the number of collection units used during the previous Census exceeds the target number of districts, or if any of the units contained more than the upper bound of the target range (i.e., 400 dwellings) during the previous Census, the CTP is redistricted to keep the number and size of districts within reasonable limits. Otherwise, they are retained. (The 1991 Census convention requires, additionally, that the lower bound be exceeded to reduce the number of very small districts.) Their shape may be altered by manual intervention, however, if there has been a change in the CTP boundary since the previous Census. A final manual check is needed to verify if the magnitude of the change to the previous boundaries is sufficiently significant to warrant redistricting.

b) DISAGGREGATION

The CTP is subdivided into T approximately equally sized parts by overlaying circular or rectangular grids that have been adapted to cover the entire area of the Bounding Box of the CTP -- depending upon the ratio of the major and minor axes of the Bounding Box (i.e., rectangular grid if < 0.80) and the ratio of the area of the CTP to its Bounding Box (i.e., rectangular grid if < 0.80). In this first data availability case, DISAGGREGATION is not performed using seed methods. Instead, rectangular grids and sectors with rings are used. This is because there is no information available about the location of blocks, streets or dwellings (below the level of the CTP), and seed-based DISAGGREGATION can not be employed without such data. Thus, only the bottom two rows of Table 4.1 are used to select a DISAGGREGATION method in this first case.

The structure of Table 4.1 is based on the relation between the ratio of the area of the CTP and the area of the Bounding Box or "index of area covered" (that is similar to the notion of "percent area covered") and the ratio of the length of the minor axis to the length of the major axis of the Bounding Box (i.e., an "elongation index").

Table 4.1 Table for Selecting Methods for Districting

ELONGATION INDEX Ratio of Minor Axis to Major Axis of the Bounding Box	INDEX OF AREA COVERED (Area of the CTP versus the Bounding Box)		
	< 0.40	0.40 - 0.80	> 0. 0
< 0.40	Random Seeds	Extrema-Based Seeds	Regular Seeds
0.40 to 0.80	Unidirectional Grids	Bidirectional Grids	Sectors with Rings
> 0.80	Rings with Sectors	Sectors	Rings

Circular grids have their origin at the centre of the Bounding Box and rectangular grids at the lower left corner of the Bounding Box. Circular grids trace out sectors about their origin starting either at a point on the Bounding Box due north of the origin or at the centroid of the block with the largest value for the characteristic being employed for the districting. Rectangular grids move from the left side of the Bounding Box to the right side and from the bottom to the top.

Ranges on the two ratios have been arbitrarily set to <0.40, 0.40 - 0.80, and >0.80 which effectively weights the use of the methods (assuming a uniform distribution on each set of index values) by dividing them (in the general case) into three probability classes, 16%, 8% and 4%, and moves the midpoint of the middle range from 0.5 to 0.6 which allows for increased utilization of preferred (i.e., more successful) methods. These ranges are variable and could be tied to a feedback or learning mechanism in the future as part of the extension of the model to an expert system.

Allocation of the different geometric templates or districting methods to the cells of this table can be arbitrary and subsequently re-allocated on the basis of empirical results. The allocation shown in Table 4.1 is based on common sense tradeoffs between the geometric properties associated with the types of grid (e.g., rectangular grid cells can be square or long and narrow and, therefore, are more adaptable to lower minor/major axis length ratios, while rings are circular and would be more effective for high minor/major axis length ratios and CTP area to Bounding Box area ratios) or on the relocatability of the various types of seeds (e.g., regular seeds are fixed relative to a given Bounding Box while random seeds are highly relocatable and, therefore, are more adaptive to cases with low ratio values for the CTP/Bounding Box areas).

(As is often done in the case of decision support systems, statistics can be kept on the relative success rate of each method for varying conditions and the allocation of the methods to the cells of the table adjusted accordingly. Alternatively, the ranges of the index values associated with each row and column class can be adjusted on the basis of production experience.)

In this data availability CASE, the expected mean area per district, a, is set to the target value, V, for the CTP. The actual area, a_i, generated from the overlay procedure for the grids just described is calculated together with the deviations from the mean area, d_i, for each district, i. (That is, $d_i = a_i - a$; $i=1,..,T$.) The evaluation function accepts the districting if the sum of the absolute deviations, D, is less than 10% of the total area of the Census Tract Part. (That is, $D = \Sigma\{|d_i|\}$ or $D = \Sigma\{|a_i - a|\}$; $i = 1,..,T$ is less than $0.1 * A$.)

In the interests of providing a complete theoretical framework for all four data availability cases, this methodology has been provided here. Implementation of this methodology involves the application of standard functions within the ARC/Info Geographical Information System (namely the generation of grid coverages and overlaying those coverages on the CTP limits). However, for pragmatic reasons, and since the feasibility is not in doubt, full implementation and testing has been postponed to post-doctoral activities.

c) AGGREGATION

As mentioned earlier, without the location of block centroids, AGGREGATION cannot be performed in this case.

2. CASE # 2 - THE CENTROIDS FOR BLOCKS ARE ALSO AVAILABLE

For this data availability case, the target value, V, for the districting process is calculated by dividing the total number of blocks in the CTP, N, by the target number of districts, T, (which is supplied to the districting program by the districting specialists and is based on a combination of judgemental and local knowledge factors). (That is, V = N / T.)

a) Check Previous Collection Units

If the boundaries of the previous collection units are unavailable, the same check is made as was indicated in 1a) above. If the previous districts are still acceptable (again, according to the 1986 Census convention), they are reused. If the previous districts are unacceptable, new collection units are created using either the DISAGGREGATION or the AGGREGATION ASSIGNS described below.

If the boundaries of the previous collection units are also available the previous boundaries are used to group the set of block centroids into collection unit districts.

If the previous districts are still acceptable, (using the same evaluation function is used as for the DISAGGREGATION process described immediately below), they are reused, otherwise new collection units are created using the DISAGGREGATION or AGGREGATION processes described in the next two sections.

b) DISAGGREGATION

The CTP is subdivided into T approximately equal parts by a circular or rectangular grid that, again, has been adapted to cover the entire area of the Bounding Box of the

CTP. The cells of the grids are used to ASSIGN constituent **block** centroids to one of the T districts. That is, **blocks** are allocated to the cell in which their centroid lies. (Although these geometric decompositions could be implemented as a point-in-polygon process, it is, in fact, more efficient to group or sort the data on the basis of the location of the **block** centroids relative to the calculated "break-points" in coordinate values.)

Alternatively, a set of T seeds (**regular** - that are related to the shape of the Bounding Box, or **random** or **extrema-based** - that have been selected to fall within the CTP) is used to make the assignments based on **proximity**. The number of centroids assigned to each seed in the case of DISAGGREGATION is, therefore, variable.

The **expected mean number of blocks, n,** is calculated by dividing the **total number of blocks** in the CTP, N, by the **target number of districts,** T. The actual number of **blocks,** n_i , in each district, i, is determined together with the deviation from the mean number of **blocks,** d_i for that district. (That is, $d_i = n_i - n$; for $i = 1,..,T$.) The evaluation function accepts the districting if the sum of the absolute deviations, D, is less than the target number of districts, T. (That is, D must be less than T where $D = \Sigma\{|d_i|\}$ or $D = \Sigma\{|n_i - n|\}$; $i = 1,..,T$.)

This means that a districting is only accepted if, on average, it has a deviation less than or equal to one **block** per district. Because **blocks** are discrete individual units (i.e., each counts as 1 unit), and because the target **number of blocks** per district can accommodate all the **blocks**, no special allowance is needed for a "residual" or remainder district with this approach.

Selection of DISAGGREGATION methods, in the absence of statistical and other cartographic data (such as the street network), is made using Table 4.1 (which, at the present time, requires some human intervention to invoke/execute the methods that have been identified by the automated derivation of the index values since the programs were originally designed for data availability cases 4i and 4ii).

The **second** and **third** ASSIGN methods are selected from the same table by re-positioning the row and column indicators to the numerically nearest column and then row class respectively. For example, if the **index of area covered** was 0.55 and the **elongation index** was 0.65, the selections would be as follows:

1. Bidirectional grids (0.40 - 0.80; 0.40 - 0.80):
2. Unidirectional grids (0.40 - 0.80; < 0.40) because 0.55 < 0.60, the midpoint of the middle column class; and
3. Sectors (>0.80; 0.40 - 0.80) since 0.65 > 0.60, the midpoint of the middle row class.

Alternatively, if the index of area covered was 0.85 and the elongation index was 0.35, the selections would proceed as follows:

1. Regular seeds (< 0.40; >0.80);
2. Extrema-based seeds (< 0.40; 0.40 - 0.80) because 0.85 is closer to the range 0.40 - 0.80 than to the range < 0.40); and
3. Sectors with rings (0.40 - 0.80; >0.80) because 0.35 is closer to the range 0.40 - 0.80 than to the range >0.80.

Thus, if an outer row or column range is chosen, the subsequent range chosen moves inward on the table. If an inner class is chosen, the subsequent choice moves outwards in the direction of the index value relative to the midpoint of the inner class. This has the net effect of collapsing the table to two ranges (0 - inner class midpoint; > inner class midpoint).

c) AGGREGATION

The cells of the circular or rectangular grids are distorted or bent to compose new districts by capturing the desired number of block centroids (±1). Collinearity of centroids identified by the sweeping action of a side of a circular or rectangular grid cell requires arbitrary assignment between cells, usually on a 'first come, first serve' basis. The same method selection table (i.e., Table 4.1) is used for DISAGGREGATION.

For the seed based methods, block centroids are assigned to the nearest seed on an iterative basis that ensures an even distribution of centroids to seeds.

Again, the evaluation function accepts the districting if the sum of absolute deviations of block counts per district is less than the target number of districts.

Thus, the availability of the block centroids permits the use of the nine grid and seed based methods for both the DISAGGREGATION and AGGREGATION processes.

The method selection procedure and evaluation function are the same for both processes. The difference between these two ASSIGN processes is that the DISAGGREGATION employs SELECTION based on the position of the block centroids relative to fixed grids and seeds, while AGGREGATION involves GROUPING those centroids relative to grids and seeds whose extent of coverage varies depending on the distribution of the block centroids.

3. CASE # 3 - THE PERIMETRE AND/OR AREA OF THE BLOCKS ARE ALSO AVAILABLE

In this data availability CASE, the perimetre and/or area values for each block are either supplied along with the block centroids or are calculated from the linear graph (or digital map) explicitly defining the blocks. Sides of blocks which are not streets (and hence unlikely to be inhabited) are excluded from the calculation of this (pseudo) perimetre.

To reduce the complexity of the block grouping problem, certain blocks are first amalgamated (or 'bonded') with neighbours if settlement on them is unlikely (e.g., traffic islands, medians and cloverleafs which can be identified from the Area Master Files by selecting certain feature codes or by identifying blocks containing no blockface centroids on the database for the previous Census and having a small area).

The target value for the perimetre option, P, is calculated by dividing the total length of the street network for a given CTP, L, (counting each internal segment twice -- once for each blockface that must be enumerated -- excluding non-street sides of blocks) by the supplied target number of districts, T. (That is, $P = L / T$.)

Alternatively, the target value for the surface area option, S, is calculated by dividing the total area of the CTP, A, by the supplied target number of districts, T. (That is, $S = A/T$.)

In general, the perimetre option is more useful to collection unit districting than the area option since it better reflects the effort required to visit each household. There are, however, numerous applications that require the entire area to be covered (e.g., search and rescue operations) and therefore the area option is a useful addition to the tool kit.

a) Check Previous Collection Units

The same check is made as indicated in 2a) above as regards the number of collection units in the previous Census and the target number of districts. If the previous collection unit districts are unacceptable, new collection units are created using either the DISAGGREGATION or the AGGREGATION approaches described below.

b) DISAGGREGATION

The assignment of blocks to districts by DISAGGREGATION is identical to that employed for centroids. The 'mean' values in these two cases are set to the calculated target values. (That is $p = P$ or $s = S$, which are the target 'means' for the perimetre and surface area respectively.) The actual perimetre, p_i, or surface area, s_i, values for each block are accumulated during the districting process. The deviations, d_i, from the mean values, for each district, i, are also calculated. The evaluation function accepts the districting if the mean of the absolute deviations, D, for the individual districts is within 10% of the mean value for the CTP as a whole. (That is, D is either less than 0.1 * P or less than 0.1 * S where $D = \Sigma\{|p_i - p|\}/T$; or $D = \Sigma\{|s_i - s|\}/T$; $i = 1,..,T$.)

This means that, on average, the size of the individual districts (to the nearest decimal place) must either fall in the range 0.9P - 1.1P or the range 0.9S - 1.1S. In light of the potentially high variability in the perimetre and areas of the blocks of a CTP and the fact that the blocks are not considered divisible by this assessment, this seems to be an adequate standard for quality.

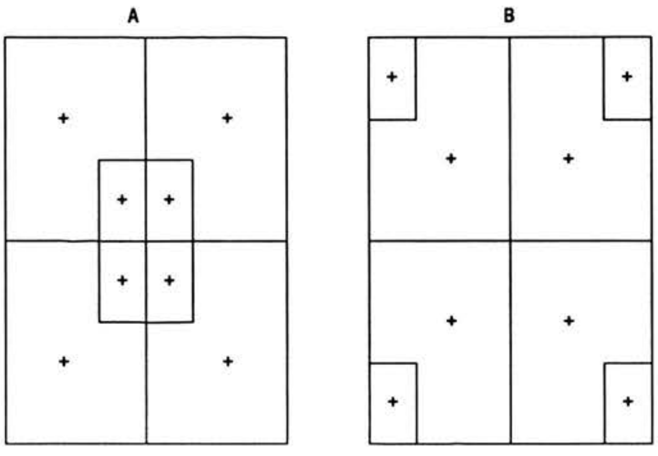
Selection of disaggregation methods, in the absence of statistical data, is made using Table 4.2 which is based on the same elongation index as Table 4.1 and a coefficient of deviation on the perimetre or area of the blocks (which serves as an indicator of the variability on the size and shape of the blocks).

o Determining the Coefficient of Deviation of "Perimeters" or Areas of Blocks

The coefficient of deviation used by the districting model to select amongst alternative districting methods (although some interaction is currently required to invoke/cause the actual execution of the identified model) is based on the standardized deviation of the "perimetres" or the areas of the blocks (that is, the absolute deviation from the mean is divided by twice the total length of the street network to account for visiting each side of the street, i.e., $CD = D/(2*L)$, or by the total area, i.e., $CD = D/A$, to give a set of values that, when summed, fall in the range 0 - 1).

These coefficients of deviation were created as surrogates to a standardized "index of dispersion" for bivariate data. The limitations of these measures as indicators of the distribution of centroids (+'s) are obvious from the examples given in Figure 4.3.

Figure 4.3 HYPOTHETICAL CONFIGURATIONS OF 8 BLOCKS



The two configurations of blocks A and B will have the same coefficient of deviation values since the perimetres and the areas are the same in both cases. However, the degree of dispersion of the centroids obviously varies considerably between the two configurations.

Nevertheless, the coefficients of deviation of either the perimetres or the areas of the blocks do give a useful measure of the relative diversity of the sizes of the blocks and hence of the complexity of districting the CTP.

A possibly more useful classification index for Table 4.2 would be a standardized multivariate index of dispersion incorporating the x and y coordinates of the locations of the block centroids together with a z value representing either the perimetre or the area of the block.

Until such an index is developed or discovered, the current surrogates will be utilized. (It is perhaps worthwhile to note at this point that one of the original reasons for developing the DISAGGREGATION models, in spite of their expected inferiority to the aggregation models, was that they are able to provide some intelligence about the dispersion of block centroids amongst the cells of the grid or seed based spatial framework.)

The allocation of the methods to the cells in Table 4.2 and the ranges on the two axes of the table are held constant from Table 4.1.

Table 4.2 Table For Selecting Methods For Districting

ELONGATION INDEX Ratio of Minor Axis to Major Axis of the Bounding Box	COEFFICIENT OF DEVIATION on the Perimetre or Area of Blocks		
	< 0.40	0.40 - 0.80	> 0.08
< 0.40	Random Seeds	Extrema-Based Seeds	Regular Seeds
0.40 to 0.80	Unidirectional Grids	Bidirectional Grids	Sectors with Rings
> 0.80	Rings with Sectors	Sectors	Rings

Thus, the distribution of the values on which the districtings are actually based begin to play a part in the selection of the districting methods. The same principle, as described above (section 2b), is used for making second and third ASSIGN method selections.

c) AGGREGATION

The cells of the circular or rectangular grid are distorted or bent (as described in Chapter 3) to capture the target perimetre length or surface area (± a threshold) based on the location of the block centroids (with some look-ahead permitted to incorporate nearby blocks whose additional workload would not exceed the target value/range whereas an intervening centroid would). Selection amongst collinear centroids is based, in part, on the value for the perimetre or area of the blocks if not all of the collinear centroids can be accommodated in the given district. That is, the collinear centroids are assigned in characteristic value order, largest to smallest, as long as the target value/range is not exceeded (by more than the tolerance).

Block centroids are assigned to the nearest seed on an iterative basis until the target value/range (perimetre or area) has been met (or slightly exceeded).

The same method selection table (i.e., Table 4.2) is utilized as for DISAGGREGATION.

Again, as for DISAGGREGATION, the evaluation function accepts the districting if the mean of the absolute deviations for the individual districts is within 10% of the mean value of the CTP as a whole.

Thus, the availability of the perimetre and area values for blocks, together (perhaps) with their explicit locations as defined by the cartographic representation of the street network, allow for more sophisticated compositions or groupings of the blocks in the CTP into collection unit districts. They also permit the method selection procedures (which are the same for both DISAGGREGATION and AGGREGATION for CASE #3 data) to consider the statistical distributions of the perimetre and area values as an indication of the complexity of the geographical distribution of the blocks (and therefore the block centroids). The evaluation functions are also the same for both DISAGGREGATION and AGGREGATION.

4. CASE # 4 - THE DWELLING DISTRIBUTION BY BLOCK IS ALSO AVAILABLE

The dwelling counts can be used alone (CASE 4i) or in combination with factors related to the perimetre and area of the blocks and the work achievement estimates and rates (CASE 4ii).

i) Only the Dwelling Counts by Block are used.

When available, a priori target values, V, (e.g., 375 dwellings) are used to determine the target number of districts, T, by dividing the total number of dwellings, H, in the CTP by the target value and rounding up to the next integer value if the result has a non-zero remainder. (That is, $T = H / V$ or $T = H / V + 1$ if $T * V$ is not equal to H for integer values T, H, and V.)

a) Check Previous Collection Units

The previous collection unit districting is re-used if the actual number of districts is less than or equal to the target number of districts and no district exceeds the upper bound of the target range of dwellings per district (e.g., 400 dwellings). In theory, this assessment can be made based on either the actual or estimated total dwelling counts for each district if the boundaries are not available. In practice, they are determined from the total dwelling counts calculated by accumulating the counts associated with the block centroids using a point-in-polygon procedure since the boundaries are available for this data availability case. Otherwise redistricting takes place.

b) DISAGGREGATION

The assignment of blocks to districts is identical to that employed for centroids except that the accumulated characteristic values are dwelling counts rather than the number of blocks. The actual counts, h_i, are accumulated during the districting process. The evaluation function accepts the districting if the mean of the absolute "deviations" is within 10% of the mean value or supplied target value (i.e., D must be less than or equal to 0.1 * V where $D = \Sigma\{| d_i |\}/M$; i=1,..,M and M is the target number of districts minus 1 to allow for the residual district). The deviations, d_i, are calculated as follows:

- between the lower and upper bound, $d_i = 0$ (e.g., between 350 and 400 dwellings);
- below the lower bound, d_i = (lower bound - district dwelling count); d_i is set to zero for the residual (i.e., smallest valued) district; and
- above the upper bound, d_i = (district dwelling count - upper bound)**2 (e.g., above 400 dwellings).

The decision to utilize exponential penalty rates above the upper bound is based on the potential negative impact of oversized districts on the quality and the timeliness of the Census collection activity and was finalized after discussions with representatives from the area responsible for Census data collection.

The choice of DISAGGREGATION SELECTION methods is made using the Table 4.3, which is based on the same coefficient of deviation on the perimetre or area of blocks as Table 4.2 and on a standardized 'special' coefficient of variation, SCV, on the number of dwellings.

The purpose of the SCV is to give a measure of the variability of the characteristics being accumulated (e.g., dwelling counts or elapsed time values) over the set of blocks whose variability in shape is measured by the coefficient of deviation, CD, of their "perimetres" or areas. The two measures, the SCV and the CD, are used in combination to select an appropriate districting methods from Table 4.3.

The standardized 'special' coefficient of variation, or SCV is calculated as follows:

1. the characteristic value c_i for each block (i.e., dwelling count values h_i or PFOM values f_i) is weighted by division by the area of the block, a_i to give

a density value, r_i;

2. the mean value, r, of r_i is computed;
3. the standard deviation, SD, is calculated;
4. the coefficient of variation, CV, is determined (Gaille has observed that "the coefficient of variation measures are highly affected by the value of the mean and share the distributional problems of the standard deviation [Gaille, 1984, p. 226]); and
5. the result is standardized by dividing by the square root of the number of blocks, N, to give standardized 'special' coefficient of variation, SCV, values in the range 0 - 1.

FORMULAE:

(1) $r_i = c_i / a_i$

(2) $r = \Sigma \{r_i\} / N$ for $i=1,..,N$

(3) SD = $(\Sigma\{(r_i - r)**2.0\}/N)**0.5$

(4) CV = SD / r

(5) SCV = CV/(N)**0.5

(The author acknowledges the substantial contribution of a statistician on the implementation team, Mr. S. Brockwell, in assisting with the formulation, derivation and implementation of a spatially oriented index for this element of the method selection sub-component of the model. While the empirical results documented in the next chapter reduced the need for these method selection methodologies, they remain a useful first step towards automating method selection for an eventual expert system.)

Table 4.3 Table for Selecting Methods for Districting

'SPECIAL' COEFFICIENT OF VARIATION on the Dwelling Count Value or the PFOM Value	COEFFICIENT OF DEVIATION on the Perimetre or Area of Blocks		
	< 0.40	0.40 - 0.80	> 0.80
> 0.40	Random Seeds	Extrema-Based Seeds	Regular Seeds
0.20 to 0.40	Unidirectional Grids	Bidirectional Grids	Sectors with Rings
< 0.20	Rings with Sectors	Sectors	Rings

The ranges on the row index (i.e., the SCV) were established empirically based on the results from the pilot system. [Brockwell, 1988]

Thus, the distribution of dwelling counts play a role, together with the distribution of perimetre or area (default is perimetre) values in the selection of districting methods and support even more sophisticated compositions of blocks into districts.

The same principle as described above (section 2b) is used for making second and third method selections.

c) AGGREGATION

The cells of the circular or rectangular grids are distorted or bent to capture the target number of dwellings (± a variable threshold) based on the location of the block centroids (with some look-ahead permitted). Selection amongst collinear centroids is identical to that employed for the perimetre and area options above. Assignment of block centroids to seeds is similarly identical to that employed for the perimetre and area options with one exception. That is, for extrema-based seeds, "extreme values" relate to the distribution (i.e., the largest number) of dwelling counts (which is also done for the CASE 4ii as described below) rather than to the geographic location of the block centroids themselves.

The evaluation function is identical to that described for DISAGGREGATION of dwelling counts (section 4ib).

ii) The dwelling counts are combined with the perimetre and areas and work achievement rates to form a Partial Figure of Merit ("PFOM") Index.

While the use of dwelling counts and customized methods for calculating the deviations for collection units closely models the traditional manual districting process, it does not take into consideration a number of more intuitive factors which form part of the subjective evaluation of districting under certain circumstances including:

1. total distance traversed;
2. dwelling density and potential for growth or "infill development";
3. remoteness considerations (of greatest concern in rural areas), and, where appropriate;
4. discontiguity considerations.

A multicomponent objective function based on these elements and on the distribution of the dwellings was therefore constructed to evaluate the districtings as a whole (i.e., at the level of the CTP). This assessment is termed the 'Figure of Merit', or FOM, of the districting. Some of the components (3 of 5) are also used in constructing the individual districts based on elapsed time (i.e., it guides the solution process) and this subset of the components of the Figure of Merit is termed the 'Partial Figure of Merit' or PFOM.

The key to combining the components of the FOM is a common measurement unit on which a correspondence between widely varying criteria is established. The unit selected is the "cost" or "effort needed" to complete the enumeration process as measured in elapsed minutes of time.

Figure of Merit (FOM) Components

This section describes a set of five indices and measures which, when combined, forms the overall "figure of merit" and which is used to assess the districting as a whole for any CTP. As mentioned, three of these components also act as an objective function for controlling the assignment of blocks, etc., to provisional collection units and form the 'Partial Figure of Merit'.

Each of the five terms of the expression defining the figure of merit are of the form:
weighting * units of effort * time per unit of effort.

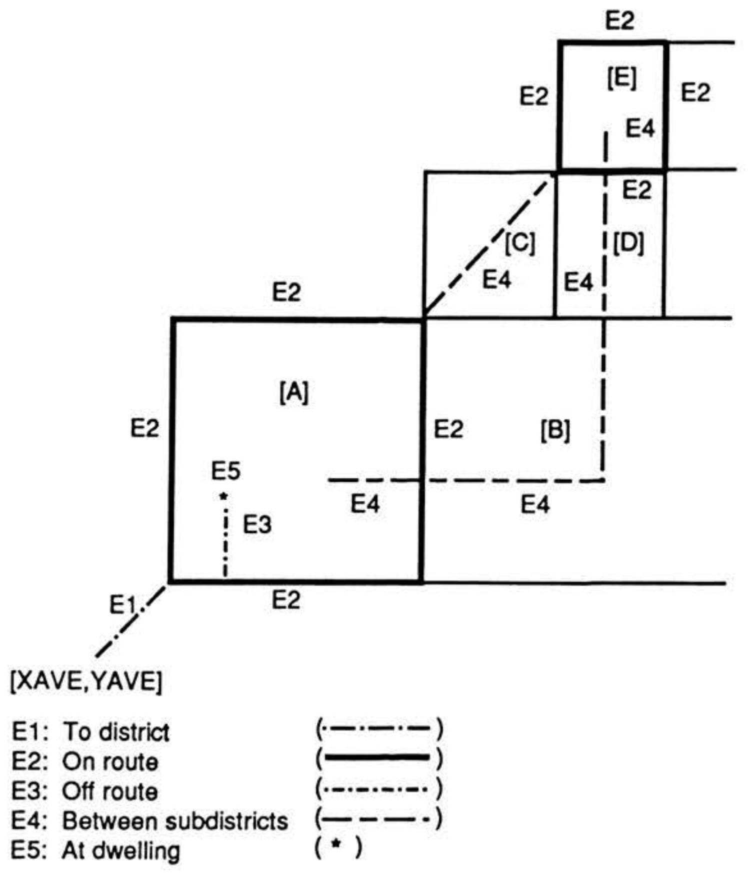
Thus the figure of merit (FOM) for a given district is defined as:

$$FOM = \Sigma\{w_i * e_i * t_i\} \text{ for } i=1,..,\text{number of components},$$

where t_i is the average time per unit of effort of type i and e_i and w_i are the effort and weighting for component i for i = 1,..,5. Since the units of time associated with each component is of equal value in the case of census collection unit districting, the value of each w_i is set to 1.0.

The relationship between each of these elements of the enumeration process is graphically illustrated in Figure 4.4.

FIGURE 4.4 An Illustration of the Elements of the FOM



The collection unit in this example consists of two subdistricts which are comprised of blocks [A] and [E].

Since the terms of the objective function allow for evaluation of all of the characteristics which result in variable workloads in the manual system, it is possible, with a computer-assisted approach, to move to a target range that combines all of the elements of time in the objective function. This range has been set at 1400 to 1600 minutes (or 23.3 to 26.6 hours) and is based both on timings taken during the 1981 Census [Statistics Canada, 1984], and on measurements made during field work at the case study site.

Each element of the Figure of Merit is considered further below in order of relative importance (rather than in the logical sequence of operations as shown in Figure 4.4 which determined the assigned index number, R1 to R5):

o **Dwelling Count Considerations (At Dwellings)**

The most critical component of the FOM is the time taken to deliver questionnaires to the individual dwellings. The number of dwellings is multiplied by the constant rate, (R5), of 3.3 minutes per dwelling [Statistics Canada, 1984] to get the total time for the district (subdistrict or block). This rate was measured during the 1981 Census of Population and Housing.

o **Block Perimetre Considerations (On Route)**

A fixed rate of travel, (R2), of 15 minutes per kilometre (i.e., 4 km/hour or 2.5 miles/hour based on empirical measurement during field work) is applied to the street length of the linear graph comprising the district (subdistrict or block) route and yields the time needed to cover the route once. To facilitate calculations and avoid costly route determination processes that would be unlikely to reflect the actual enumeration sequences (unless strictly enforced), it is assumed that all routes require a fixed rate of 'doubling back' per kilometre. This 'doubling back' is accounted for within the rate, R2. Excluded from accumulated totals, however, are block sides associated with non-road features (e.g., rivers, powerlines, etc.) as they are not typically traversed by Census Representatives.

o **"Block Dwelling Density" Considerations (Off Route)**

In physical terms, this factor, (E3), accounts for the time spent traversing the distance from the route (i.e., the road side) to the dwelling doorstep. It also brings into consideration the variable settlement densities by lengthening the distance traversed by a variable amount (up to 300% for very low densities) as defined below and thereby compensates for underestimates in the actual number of dwellings to be enumerated (e.g., due to intercensal growth).

That is, the 'off route' distances are scaled upwards if the density of the block is relatively low compared to the average for the CTP. This has the converse effect of reducing the actual number of dwellings assigned to the district since the average time for enumerating the dwellings in that block is increased.

The value of this element of the FOM is determined by multiplying a constant distance weighting factor of 10 metres per dwelling times a 'Variable Density Coefficient' (VDC) times a fixed rate of travel, (R3), (i.e., 15 minutes per kilometre) times the number of dwellings on the block, h_i, times two to allow for the time to return to the sidewalk (i.e., 10 * VDC * R3 * h_i * 2).

The VDC seeks to ensure that the size of the unit and its potential for unforeseen development are taken into consideration. The VDC consists of three factors which are each multiplied by the constant distance weighting factor (i.e., 10 metres per unit):

1. a constant 'time buffer' per dwelling, of value 1, which results in the equivalent of 10 metres of distance when multiplied by the constant distance weighting value;
2. a variable accounting for the size of the block relative to the total area for the Census Tract Part: a_i /A where a_i is the area of the i th block; and A is the area of the Census Tract Part; the value for this component falls in the range 0 - 1 and gives a distance of 0 - 10 metres when multiplied by the constant distance weighting factor; and
3. a variable accounting for the distribution of dwellings relative to the total number of dwellings for the Census Tract Part which is subtracted from 1 to obtain the desired complementary relationship: 1-h_i /H where h_i is the dwelling count for the i th block; and H is the dwelling count for the Census Tract Part. Again, this gives a value in the range of 0 - 10 metres when multiplied by the constant weighting factor.

The resulting value for the VDC ranges between 1 and 3 and the mathematical formulation for the variable density coefficient is:

FORMULA:

(6) $VDC = 1 + \{a_i / A\} + \{1 - h_i / H\}$

Thus, the range in value of the effort associated with this element of the FOM is between $10*h_i*2*R3$ and $30*h_i*2*R3$. This is the same as traversing a distance of between 10 and 30 metres, twice (once coming and once going) for each of the h_i dwellings at rate of travel R3.

o Contiguity Considerations (Between Subdistricts)

The FOM penalizes proposed districts (or workloads) that are disjoint. This is achieved by including a factor that adds a fixed rate, (R4), of 1.5 minutes per block crossed or per 'topological transit' (or an amount equivalent to an approximate distance of 150 metres per transit at walking speed R2) that must be made in order to get between the closest two blocks of each subdistrict. In addition, the Manhattan distance (i.e., the absolute value of the difference between the x and y coordinates) is added to the 'to route' distances defined in the final element of the FOM. The second term is an actual cost to move between the two subdistricts while the first is an added penalty to discourage discontiguous districts.

o Remoteness Considerations (To District)

A central point (e.g., the centre of the CTP) is viewed as the starting point for each district to be enumerated. This element of the FOM is associated with the overhead of moving from the central point to the start of the district "route". A fixed rate of travel (R1), 2 minutes per kilometre (or 30 kilometres per hour) is applied to the Manhattan distance in metres to get from the central point [XAVE,YAVE] to the start of the route (which is assumed to be the closest point for ease of calculation) to yield the time needed. The values for XAVE and YAVE are either supplied **a priori** or are set to the default value which is the centre of the Bounding Box.

The index/identifier, weightings, average time per unit rates, and effort interpretation for each of the components of the FOM are summarized in Table 4.4.

Table 4.4 FIGURE OF MERIT CONSTANTS AND RATES

COMPONENT	INDEX	w_i	t_i	e_i
i=1	R1	1.00	0.002 minutes per metre	Distance to district in metres
i=2	R2	1.00	0.015 minutes per metre	Distance on route in metres
i=3	R3	1.00	0.015 minutes per metre	Distance off route in metres
i=4	R4	1.00	1.5 minutes per transit	Distance between district parts in transits
i=5	R5	1.00	3.3 minutes per dwelling	Number of dwellings

Partial Figure of Merit Components

In assigning blocks to collection units, an objective function called the Partial Figure of Merit (PFOM) is used in CASE #4ii to build the districts by grouping blocks until the target value/range threshold (in terms of minutes of effort) is achieved or slightly exceeded.

The PFOM takes into consideration three of five elements of the complete Figure Of Merit:

- o the length of the street network that is travelled, (E2);
- o the number of dwellings to be visited, (E5); and
- o the density of dwellings, (E3).

Omitted are the two elements which are determined only after the districts have been created (i.e., the distance to the start of the collection unit, E1, and the distance between parts of the collection unit, E4). Thus, the PFOM target value/ranges represent the elapsed time (minutes) to complete the major portion of the questionnaire delivery process and encompasses a significant portion of the range of the complete FOM as depicted in Figure 4.5, which also shows the relationship of the FOM and PFOM target value/range to the dwelling target value/range.

Figure 4.5 Comparative Target Values/Ranges

```
              Dwelling Value/Ranges
                   ( 375 ± 25 )
                  ( 375 ± 6.7%)

          350          375          400

  |---------|--------------|--------------|---------|

         1400          1500          1600

               PFOM Value/Ranges
                ( 1500 ± 6.7% )

 1350               1500                  1650
               FOM Value/Ranges
                 (1500 ± 10% )
```

a) Check Previous Collection Units

Until such time as the notion of a PFOM or a "time-based" assessment of the previous collection units is better appreciated, the same check is used as above for dwelling-based districtings (i.e., 4ia). That is, the previous collection unit districting is re-used if the actual number of districts is less than or equal to the target number of districts and no district exceeds the upper bound of the target range for dwellings.

Eventually, operations supervisors and staff will be as comfortable working in units of elapsed time as they are currently with dwelling counts (together with a host of exception conditions to deal with expected growth, size of district, etc.), and it will be possible to consider assessing the previous collection unit districts in term of criteria (i.e., time) other than those by which they were initially created (i.e., dwellings).

b) DISAGGREGATION

The assignment of blocks to districts is identical to that employed for centroids. The evaluation function accepts the districting if the mean of the absolute "deviations" is within 10% of the supplied target value (i.e., D must be less than or equal to 0.1 * V where D = SUM $\{d_i\}/M$; i=1,..,M and where the "deviations", d_i, are calculated as follows:

- o between the lower and upper bound (i.e., between 1400 and 1600 minutes), $d_i = 0$;
- o below the lower bound (i.e., 1400 minutes), d_i = (lower bound - PFOM value) which is set to zero for the residual (i.e., smallest valued) district; and
- o above the upper bound (i.e., above 1600 minutes), d_i = ((1 + (PFOM value - upper bound)/4.0)^2)*4.0.

The use of an exponential penalty factor mirrors the approach taken for the dwelling-based calculation of deviations. Division by four maps the midpoint of the time-based assessment (i.e., 1500 minutes) to the midpoint of the dwelling-based assessment (i.e., 375 dwellings) and dampens the impact of the exponential penalty factor to more reasonable limits. Multiplication by four effects the reverse mapping while the addition of one (1) ensures a minimum deviation for fractions caused by division by four that, when squared, become smaller rather than larger.

Selection of the PFOM based DISAGGREGATION method is based on the same table (i.e., Table 4.3) as for dwelling-based DISAGGREGATIONS and AGGREGATIONS.

c) AGGREGATION

The method for AGGREGATION of PFOM values is the same as for the

AGGREGATION based on dwelling counts. The evaluation function is the same as for DISAGGREGATION based on PFOM values above.

The currently implemented method selection table, (Table 4.5), pre-dates Table 4.3 and is based on the same standardized 'special' coefficient of variation of the dwelling count case (4i) and on the target number of districts. The allocation of methods to the cells of the table was based on the geometric properties of the different methods. With two exceptions, the allocation is identical to those of Tables 4.1, 4.2 and 4.3. The fact that ring methods seem inappropriate for nine or more districts while rings and sectors would make little sense for less than four districts meant that these two methods are interchanged relative to their positions in the other tables.

Thus, during the initial (i.e., pilot) testing, Table 4.5, (which has, however, been reorganized in light of testing results presented in Appendix G) was tested as a mechanism for choosing districting methods for PFOM based districting. (This testing required that all methods be tried for each of the 63 CTPs in Laval to verify whether or not the mechanism was more effective than random choices. The fact that method selection using the number of districts as an index performed only slightly better than making random selections resulted in a consideration of other indices which, in turn, led to the development of decision tables 4.1., 4.2, and 4.3.)

Table 4.5 Table for Selecting Methods for Districting

'SPECIAL' COEFFICIENT OF VARIATION on the Dwelling Count Value or the PFOM Value	NUMBER OF DISTRICTS		
	2 - 4	5 - 8	9 or more
> 0.40	Random Seeds	Extrema-Based Seeds	Regular Seeds
0.20 to 0.40	Unidirectional Grids	Bidirectional Grids	Sectors with Rings
< 0.20	Rings	Sectors	Rings with Sectors

The general principles behind the allocation of the methods to the cells of the table included the following:

1. seed methods were viewed as more suitable for surfaces with higher variability while circular and rectangular grids were seen as more suitable for surfaces

with lower variability;

2. unidirectional grids (rectangular and circular) are allocated, together with the highly flexible random seed method, to cases requiring only a small (2-4) number of districts;

3. bidirectional grids and regular seeds (which closely approximate a bidirectional grid) are allocated to cases with larger numbers of districts;

4. finally, rectangular grids are allocated to the mid-range case of data variability to maximize their use (rather than circular grids) because of their relatively high level of success during prototype testing.

The comparative merits of each of the method selection approaches (as depicted in Tables 4.1, 4.2, 4.3 and 4.5) are considered in the next chapter.

Thus, the availability of work effort rates together with dwelling count distributions and the perimetre and area values for individual blocks meant that the full amount of information needed for emulating the manual process could be combined (by establishing a common unit of measure) to form, first, an objective function for forming districts, secondly, a figure of merit for the districting as a whole, and finally, an index for choosing between districting methods.

d) BLOCK OR SUBDISTRICT REGROUPING

REGROUPING is first performed by the ANNEX process at the level of subdistricts that were generated by the ASSIGN process and then by the RE-ASSIGN and ADJUST processes at the level of blocks within the entire CTP. The ADJUST process involves manual intervention and can therefore be employed in all of the data availability CASES but is generally reserved as a last (most costly) resort. As currently implemented, the ANNEX and RE-ASSIGN processes require that CASE #4 data be provided (but they can easily be extended to be used for CASE #3 data). Each REGROUPING process is now described in greater detail.

1. The ANNEX Process

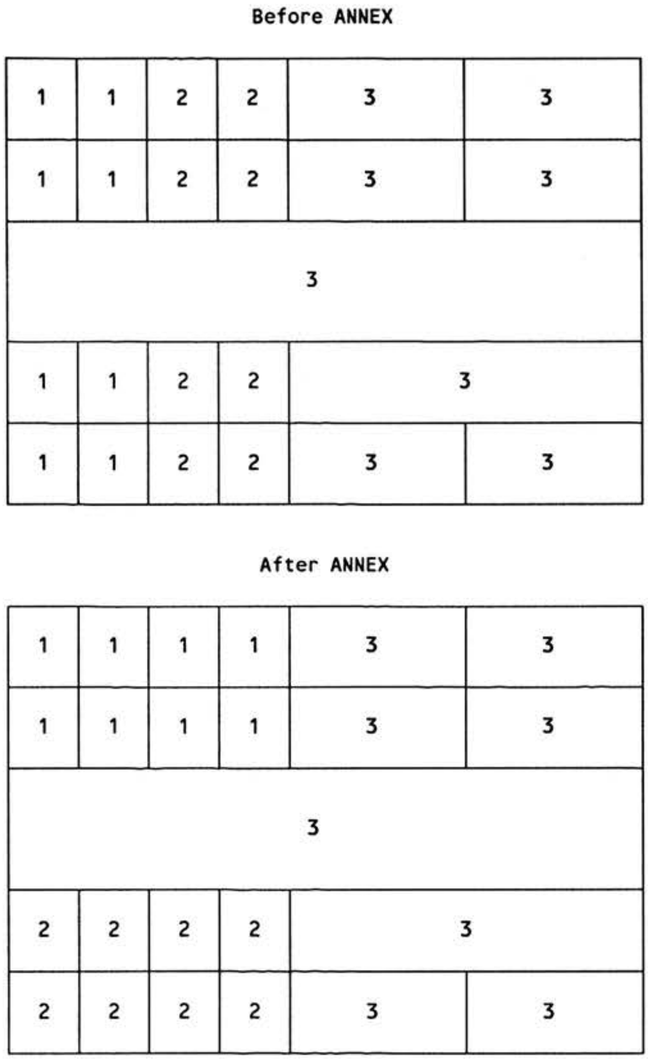
The ASSIGN process produces districts by aggregating areal units (i.e., blocks represented by their centroids) using a 'globally' applied (fixed or variable) model. The RE-ASSIGN process, on the other hand, composes districts by 'locally' aggregating neighbouring areal units (i.e., blocks with their topology which is either supplied or derived from the cartographic representation of their boundaries), step-wise about a set of seeds or "cores". The ANNEX process combines the results of 'globally' produced ASSIGNs and 'locally' re-allocates portions of districts (i.e., 'subdistricts') to contiguous neighbouring districts/subdistricts and might be considered, therefore, to work at an 'intermediate' level.

During prototype testing the ASSIGN process proved to be very quick, but led to discontiguities. The RE-ASSIGN process (described in detail in the next section) used for the prototype system ensured that there were no discontiguities, but was, consequently, more costly and led to oversized districts on occasion. The ANNEX process was developed, therefore, as an inexpensive 'add-on' to the ASSIGN process to ensure that discontiguities were eliminated -- occasionally at the expense of forming more than the target number of districts. (Similarly, the 1991 Census production system version of the RE-ASSIGN process prohibits oversized districts since the ANNEX process can be used to reduce the number of districts to a more manageable total.)

The ANNEX process takes as input the results of an ASSIGN. (At the suggestion of a member of the team implementing the production system, an automated check is made to ensure that the sum of the two smallest district/subdistricts is less than the target value since this must be true for an ANNEX to be performed. During the testing of the prototype system, such checks were performed by visual inspection.) The discontiguous elements (districts and discontiguous district parts) are all treated as 'subdistricts' and ranked according to size. Starting with the largest 'subdistrict' (in terms of the number of dwellings or the value of the PFOM), 'subdistricts' are 'annexed' (largest to smallest) by the largest neighbouring subdistricts until such time as the upper bound would be exceeded if any of the neighbouring subdistricts were added. The ANNEX is completed when the last 'subdistrict' is re-allocated. If none of the subdistricts can be combined by the ANNEX process, the districting will be unchanged and each subdistrict is evaluated as if it were a district before passing to the next ASSIGN or RE-ASSIGN step.

The utility of the ANNEX process is provided in a hypothetical example portrayed in Figure 4.6. Assume that each of the 24 blocks has a dwelling count of 50 dwellings and that the first diagram, "Before ANNEX", depicts the result from an ASSIGN by method 9, UNIDIRECTIONAL RECTANGULAR GRID. Each district contains the desired number of dwellings, 400, but districts 1 and 2 would each contain two discontiguous subdistricts. The second diagram shows how the ANNEX capacity would re-group those subdistricts into contiguous districts with the desired number of dwellings.

Figure 4.6 Hypothetical Example of the ANNEX Process



Since the ANNEX capacity operates at the intermediate level of the 'subdistrict', rather than at the level of the individual areal units (i.e., blocks), and since the topological relationships between the "subdistricts" are explicitly represented within the Geographical Information System, ARC/Info, the cost of this operation is very low. (Indeed, the cost is so low that, for the 1991 Census production system, even successful results from the ASSIGN process and previous districts are revised by the ANNEX process in the hope of reducing the number of districts below the target.)

The evaluation function accepts the results of the ANNEX process if the same criteria as the preceding ASSIGN process are satisfied.

2. The RE-ASSIGN Process

If the evaluation of the results of the ASSIGN and ANNEX processes shows that none of the three selected methods is acceptable, the RE-ASSIGN process is performed on, at most, the best two results from the ASSIGN and ANNEX processes.

Given the cost of building the block topology, it makes sense to take advantage of that structure for more than one RE-ASSIGN attempt (if necessary). On the other hand, given the computational cost of performing the RE-ASSIGN process, it was assumed (and later demonstrated) not to be cost-effective to perform the RE-ASSIGN process on all three unsuccessful results from the ASSIGN process. (This judgement had as much to do with failure rates as with computation costs.)

- RE-ASSIGN Seed Selection

  The first step in the RE-ASSIGN process is to select a number of core blocks equivalent to the target number of districts from the rejected collection units of the most favourable of the three unacceptable districtings from the ASSIGN and ANNEX steps. The characteristic-weighted centres of gravity of the rejected collection units or seeds (C1, C2, and C3 in the example in Figure 4.7) are used to select a "core" block (A, B, and C in Figure 4.7) for each seed based on proximity (using Pythagorean distance). Each block may be the core for at most one seed. (Identifying cores in this manner takes direct advantage of the information about the distributions of blocks and the characteristic values being used for the specific districting. It also takes direct advantage of the information gained about the localization or clustering of the block centroids during the ASSIGN process.)

- Topological RE-ASSIGN of Non-Shared Blocks

  Re-assignment of blocks proceeds iteratively outwards, in terms of both topological distance (measured as 'topological transits') and Pythagorean distance (if the characteristic values are equivalent), from each core until:

- all blocks are assigned;
- the target value is reached or exceeded (by less than the tolerance value); or
- blocks are found to be neighbours of, or "share", two or more of the "growing" districts.

Immediately neighbouring blocks (i.e., blocks sharing at least one side or one node) are RE-ASSIGNED in order (largest to smallest value for the aggregation characteristic).

o RE-ASSIGNMENT of Shared Blocks

Blocks that (in terms of the number of 'topological transits' required to reach the core blocks) are equally adjacent to, or 'share' a side or node with, two or more expanding districts are 're-assigned' last (to encourage the re-assignment of peripheral blocks to the nearest cores). In order to increase the ease of adding blocks to core based districts that are nearly full, shared blocks are allocated, in descending order, to cores able to accommodate their workload in accordance with the following rules of precedence:

- the core based districts with the fewest degrees of freedom (i.e., the fewest shared blocks) are given priority (to prevent those districts from becoming prematurely isolated, i.e., with no unassigned shared blocks),
- core based districts with the lowest accumulated total are assigned shared blocks until they no longer have the lowest accumulated total or until a neighbouring district has fewer shared blocks (to level the load amongst districts), and finally
- if all other factors are equivalent, the core based district with the smallest land area is given the next block (to balance the size/density of the districts).

A hypothetical example of how the RE-ASSIGN process functions is portrayed in Figure 4.7 which takes the same input as Figure 4.6 (i.e., the "Before ANNEX" diagram). The first diagram of Figure 4.7 shows how the core blocks (displayed with a '+' and labelled A, B, and C) are selected on the basis of proximity to the characteristic weighted centres of gravity (or "seeds") of the block centroids of the rejected collection units (displayed with an '*' and labelled C1, C2, and C3 in the latter case) for the districts produced by the

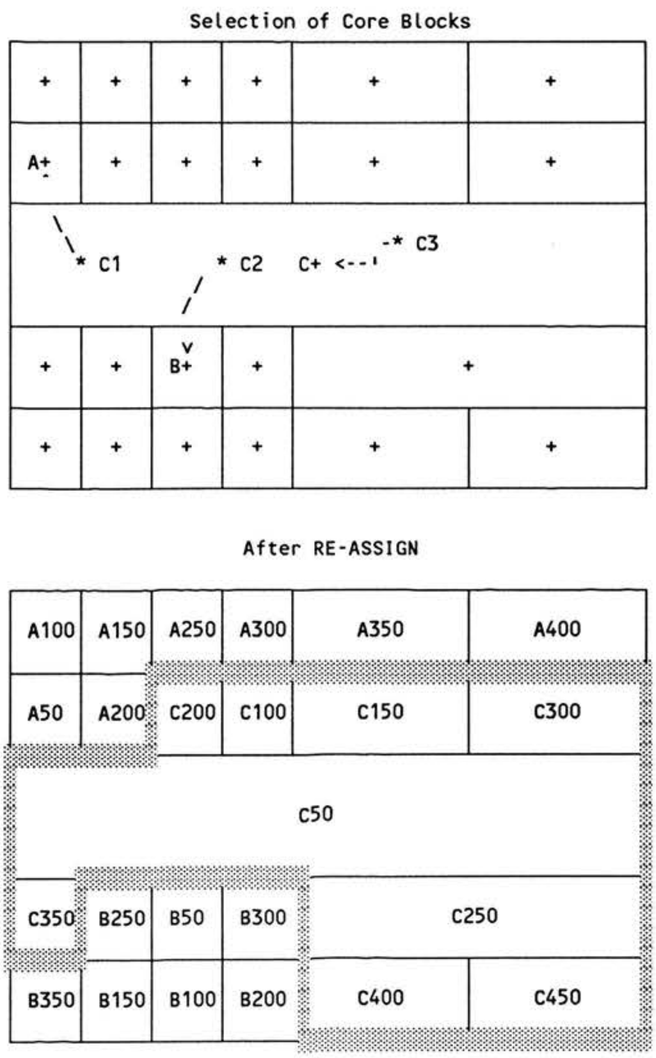
ASSIGN (from the "Before ANNEX" component of Figure 4.6).

The second diagram of Figure 4.7 shows the block re-groupings from the RE-ASSIGN process. The re-assignment sequence is indicated by the letter of the core block followed by the accumulated total number of dwellings. (That is, the sequence A50, A100, A150 indicates the first three blocks allocated to district A.)

The resulting districts (which are outlined with a bold line) are contiguous but range in size from 350 dwellings for district A to 450 dwellings for district C.

The evaluation function accepts the results of the RE-ASSIGN process based on the same criteria as for the previous ASSIGN process.

Figure 4.7 Hypothetical Example of the RE-ASSIGN Process



3. The ADJUST Process

There are two kinds of ADJUST processes.

1. Adjusting the criteria (i.e., allowing "exemptions").

   For example, the number of districts might be allowed to exceed the target, the size of a particular district might be allowed to exceed the upper bound, or two subdistricts might be considered adequately contiguous if they share a common node or street intersection.

2. Adjusting the district boundaries (e.g., swapping blocks).

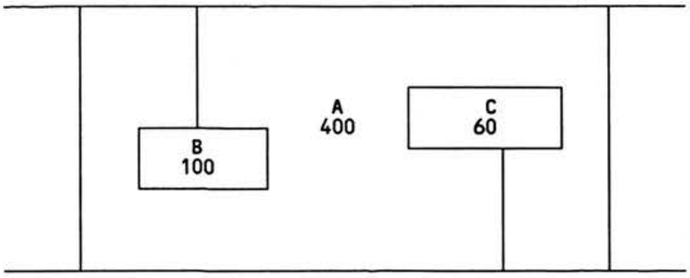
   If none of the ASSIGN, ANNEX or RE-ASSIGN processes are judged successful by the automated assessment mechanism or are rejected by a supervisor, human intervention is employed to ADJUST the boundaries of the best of the result from unsuccessful automated districtings until an acceptable districting is created.

   It should also be noted that such intervention is possible at any time during the ASSIGN, ANNEX or RE-ASSIGN processes. (For example, it might be most cost-effective in certain deadlock cases -- as described below -- to invoke the ADJUST process after the completion of the first ASSIGN-ANNEX process.)

   o Deadlock Resolution

     Generating "optimal" collection units and resolving allocation situations that are deadlocked (e.g., consider an individual block, 'A' in Figure 4.8, with a dwelling count that is equal to the upper bound of the target range and that also contains several small blocks, 'B' and 'C', with additional dwellings) can quickly become cost prohibitive if a totally computerized solution is over-emphasized.

Figure 4.8 Deadlock Case Example



The current model has been designed to take advantage of the expertise of the manual districting staff to resolve such deadlocks and/or "optimize" allocations using batch or interactive sessions to allocate or re-allocate blocks to districts in difficult areas using the ADJUST module.

The results from an ADJUST process can be assessed using the same evaluation function as the step from which it was invoked (or called for). However, in practical terms, the ADJUST is only complete when the districting specialist has decided that the results are operationally adequate. Thus, a formal evaluation may not take place.

4.8 SUMMARY

The set of districting models and processes described in this chapter provide a powerful and flexible set of tools for responding to varying conditions and data availability.

The complexity and sophistication of the models increase with the amount of available information.

The focus for comparative testing is, however, on the last data availability CASE (4i and 4ii). From an operational perspective, the expectation is that, initially, interest will focus on the dwelling based approach to districting since comparison with results from traditional manual approaches will be more straightforward if based on dwelling counts. This is expected to be superseded by the elapsed time based approach once there is adequate confidence in the autodistricting capacity.

This section summarizes the decision trees and options related to the selection of: Tables for Method Selection; the criteria used to build each table; the origins and transformations of each underlying grid or seed based model; and the application of variable evaluation functions appropriate for the available data.

The relationship between the availability of data, the selected approach and the number of the Method Selection Table are summarized in Table 4.6.

Table 4.6 METHOD SELECTION - TABLE USED

Digital Data Availability (Cumulative)	APPROACH SELECTED	
	DISAGGREGATION	AGGREGATION
Census Tract Part Boundaries	Lower part of Table 4.1	Not Applicable
Block Centroids	Table 4.1	Table 4.1
Street Network Pattern	Table 4.2	Table 4.2
Collectives and Blockface Dwelling Counts	Table 4.3 or Table 4.5	Table 4.3 or Table 4.5
Partial Figure of Merit Rates	Table 4.3 or Table 4.5	Table 4.3 or Table 4.5

The criteria employed in constructing the indices for each of the Method Selection Tables is summarized in Table 4.7.

Table 4.7 METHOD SELECTION - CRITERIA USED FOR INDICES

Digital Data Availability (Cumulative)	APPROACH SELECTED	
	DISAGGREGATION	AGGREGATION
GEOGRAPHIC DATA Census Tract Part Boundaries	$\frac{\text{Area of CTP}}{\text{Area of Box*}}$ and $\frac{\text{Minor Axis Length*}}{\text{Major Axis Length*}}$	Not Applicable
CARTOGRAPHIC DATA Block Centroids	$\frac{\text{Area of CTP}}{\text{Area of Box}}$ and $\frac{\text{Minor Axis Length}}{\text{Major Axis Length}}$	$\frac{\text{Area of CTP}}{\text{Area of Box}}$ and $\frac{\text{Minor Axis Length}}{\text{Major Axis Length}}$
CARTOGRAPHIC DATA Street Network Pattern	Deviation in Block Perimeter or Area and $\frac{\text{Minor Axis Length}}{\text{Major Axis Length}}$	Deviation in Block Perimeter or Area and $\frac{\text{Minor Axis Length}}{\text{Major Axis Length}}$
STATISTICAL DATA Sub-block Collection Unit and Blockface Dwelling Counts	Deviation in Block Perimeter or Area and Variation in Dwelling Counts	Deviation in Block Perimeter or Area and Variation in Dwelling Counts
OPERATING RATES Partial Figure of Merit Rates	Deviation in Block Perimeter or Area and Variation in PFOM Values ---- or ----- Variation in PFOM Values and Number of Districts	Deviation in Block Perimeter or Area and Variation in PFOM Values ----- or ------ Variation in PFOM Values and Number of Districts

* The 'area', 'minor axis' and 'major axis' refer to the rectangle formed by the minimum and maximum coordinates or 'Bounding Box' of the Census Tract Part.

The relationship between the location of the grid and seed assignment models and approach is summarized in Table 4.8.

Table 4.8 BASIS FOR LOCATING GRIDS OR SEEDS

Method	APPROACH	
	DISAGGREGATION	AGGREGATION
GRIDS 1. Rectangular a. Unidirectional b. Bidirectional	Fixed - based on the location of the Bounding Box and the number of districts and the direction of the major axis	Variable - based on the location of the Bounding Box, the block centroids and the number of districts and the direction of the major axis
2. Circular a. Unidirectional i. Rings ii. Sectors b. Bidirectional i. Rings with Sectors ii. Sectors with Rings	Fixed - based on the centre and size of the Bounding Box and the number of districts	Variable - based on the centre and the size of the Bounding Box; the location of the block centroids and their counts; and the number of districts
SEEDS 1. Regular	Fixed - based on the location of the Bounding Box and the number of districts	Fixed - based on the location of the Bounding Box and the number of districts
2. Random	Variable - based on the location of the Bounding Box and the number of districts	Variable -based on the location of the Bounding Box and the number of districts
3. Extrema-based	Variable - based on the location or the value of the block centroids and the number of districts	Variable - based on the location or the value of the block centroids and the number of districts

The tests that are used to evaluate the districtings for each data availability CASE are summarized in Table 4.9.

Table 4.9 EVALUATION FUNCTION - OPTIONS AVAILABLE

Data Available (Additive)	APPROACH SELECTED			
	FILTERING	PREVIOUS EA UNITS	DISAGGREGATION	AGGREGATION
Census Tract Part Boundaries	Does Not Apply	Test 2 Previous EA Based	Test 4 Area-Based	Does Not Apply
Block Centroids	Does Not Apply	Test 2 Previous EA Based	Test 5 Block-Based	Test 5 Block-Based
Street Network Pattern	Does Not Apply	Test 2 Previous EA Based	Test 6 Perimetre or Area Based	Test 6 Perimetre or Area Based
Sub-block Collection Unit and Blockface Dwelling Counts	Test 1 Sub-Blockface, Blockface, and Block Units	Test 3 Previous EA or Centroid Dwelling Based	Test 7 Perimeter, Area or Dwelling Based	Test 7 Perimeter, Area or Dwelling Based
Partial Figure Of Merit Rates	Test 1 Sub-Blockface, Blockface, and Block Units	Test 3 Previous EA or Centroid Dwelling Based	Test 8 Perimeter, Area, Dwelling or PFOM Based	Test 8 Perimeter, Area, Dwelling or PFOM Based

Test 1: Individual districts are within target range (except sub-blockface collection units which may be smaller)

Test 2: Previous districts are less than upper bound and the number of districts does not exceed the target number.

Test 3: Re-compiled previous district counts are less than the upper bound and the number of districts does not exceed the target after filtering.

Test 4: The sum of deviations in the area of districts is less than 10% of the total area.

Test 5: The sum of deviations in the block count is less than the number of districts.

Test 6: The sum of deviations in perimeter, or area of the district is less than 10% of the mean value of the CTP.

Test 7: The sum of deviations in perimeter, area, or dwelling count values is less than 10% of the mean value of the CTP.

Test 8: The sum of deviations in perimeter, area, dwelling count, or PFOM values is less than 10% of the CTP mean value.

4.9 CONCLUSIONS

This chapter has provided an overview and description of the modelling strategies and of the implemented districting model including the main components and their interrelationships in the usual process flow or mode of operation. The components of the model were described in detail in the context of four data availability CASES or scenarios.

Based on a limited number of clearly identified assumptions, a methodology has been developed and implemented as a computer-based system for solving the census collection unit districting problem (Chapter 3).

The parameters of the model were initially established inductively and were further refined through case study testing. A discussion of the testing strategies and the results from a case study are presented in the next chapter.

CHAPTER 5

CASE STUDY TESTING AND ANALYSIS OF RESULTS

5.1 INTRODUCTION

This chapter describes the test framework, first for verifying the proper functioning of the model, and then for measuring the model's performance through a series of case study tests which increase in volume, complexity and scope.

The testing was conducted in three stages:

1. validation of the proper functioning of the implemented model on 6 controlled data sets or "test patterns";
2. assessment of the feasibility of the validated model on 4 actual test areas; and
3. assessment of the viability (performance) of the proven model on 61 Census Tracts (forming 63 CTPs due to CTs split by FEDs).

The results of these tests are presented and analysed from both an operational and a theoretical perspective. Limitations on the selection of the case study areas, the nature of the tests, the comparison of results from manual versus automated processes, the restructuring techniques to enhance the utility of the Area Master Files (e.g., 'block bonding' methods), and the utility/applicability of the new autodistricting capacity are also discussed.

5.2 METHODS OF ASSESSMENT

In keeping with traditional assessment methods, a PASS/FAIL assessment for disaggregation and aggregation by dwellings was initially made only on the basis of the acceptability of the dwelling count distributions. To simplify the evaluation and the process of technology assimilation, information on the distribution of blocks, perimetres, surface area and the elapsed time based partial figure of merit were produced by the system but were not formally used as goals in the early stages of testing the model.

5.2.1 PASS/FAIL Evaluation Function Based On Dwelling Counts

With autodistricting, fourteen of the twenty characteristics (see Table 5.16) are guaranteed as a matter of procedure. Thus, only the six remaining characteristics need be formally considered:

1. Facilitate Accessibility;
2. Respect Enumerator Workload Limits;
3. Respect Contiguity;
4. Minimize Route Lengths;
5. Minimize Route Start Distances; and
6. Strive For Compact Shapes.

Of these, respecting enumerator workload limits was the first criterion to be handled by automatic dwelling-based assessment since it represents one of the most important indicators of quality in the traditional manual approach.

Four of the five remaining characteristics are also included explicitly as part of the Figure of Merit based assessment, namely, facilitate accessibility, respect contiguity, minimize route lengths and minimize route start distances. While the fifth characteristic, strive for compact shapes, is not an explicit component of the Figure of Merit, the autodistricting tool kit considers this element indirectly in a number of ways (see Section 5.9 for further details).

The enumerator "workload" is expressed in terms of dwellings to be visited and this research developed a special index for measuring deviations from the calculated target number of dwellings per district that takes into account the particular needs and values of the collection operations for a census. The evaluation function, which sums these special deviations and determines if the results are within allowed tolerances, was designed by the author and finalized after extensive consultation with districting specialists and end users.

The elements of this special index are described in detail in Chapter 4. All values within the target ranges and the residual district are assigned a deviation value of zero. The sum of the squared deviation values above the upper bound is added to the sum of the linear deviations below the lower bound to give a total deviation value.

A final assessment is made on each districting, in turn, to see if the average deviation score (the total deviation divided by the target number of districts) for the Census Tract Part is within the allowable Target Average Deviation (i.e., 10% of the target value or 37.5 dwellings per district in urban areas) for that CT Part.

For the purposes of calculating a final deviation score, the discontiguous parts (or subdistricts) are considered to be separate small districts.

However, this method of calculating the final deviation score was implemented, on the basis of experience and increased understanding of the requirements as perceived by the end users, between the second and third stage in the testing process and was not, therefore, considered in the earlier testing stages. For the two earlier testing stages, the districtings with discontiguous parts (or subdistricts) were considered acceptable. Thus, the model was viewed as operating at the level of the "workload", rather than at the level of the collection unit itself since actual "workloads" assigned to individual Census Representatives are occasionally made up of several small, discontiguous districts.

For the 1986 Census, for example, 297 manually generated collection units were reduced to 268 "workload" assignments for the CSD of Laval.

In addition to the dwelling based assessment, during the first testing stage, a second assessment was made for each of the other characteristics to verify the proper functioning of the districting model. These assessments were based on the distribution of the given characteristic -- block count, perimetre, surface area and the elapsed time (derived from the partial figure of merit calculation). Details of each of these unique evaluation functions were provided in Chapter 4 and are summarized below in order to clarify their relation to the PASS/FAIL evaluation.

5.2.2 PASS/FAIL Evaluation Function Based on Figure of Merit

For the third testing stage (and to a limited extent during the second stage), the enumerator effort is also expressed in terms of the number of minutes required to complete the component elements in the figure of merit function described in the previous chapter.

The total deviation for the districting is taken as the sum (after scaling) of the squared deviations above the upper bound and the linear deviations below the lower bound. All values falling within the target range and the residual district being assigned a deviation value of zero. Again, the final assessment is made on each districting to see if the average deviation score for the Census Tract Part is within the 10% of the Average Deviation for the target score for the Census Tract Part (i.e., 150 minutes per district in urban areas).

As above, during the third stage of testing, the discontiguous parts or subdistricts are considered to be separate districts for calculating the PFOM deviation score (i.e., discontiguities are not permitted).

5.2.3 PASS/FAIL Evaluations Based On Block Count, Perimetre And Area

During the first testing stage, the block count, perimetre and surface area based approaches were also tested (but not evaluated). Unlike the methods of evaluation based on dwelling counts and PFOM values, the target number of districts for these three approaches cannot readily be derived from the distribution of the characteristic over the surface (i.e., the total value for the characteristic divided by the characteristic target value) and it must be supplied as an input by the user to the autodistricting program. Therefore, the heuristics used to evaluate the success of these types of autodistricting vary somewhat from those described for dwelling and PFOM approaches.

For block, perimetre and area based approaches, the target characteristic value for each district is determined by dividing the total count (e.g., the number of blocks, perimetre, or surface area of the CT Part) by the supplied target number of districts.

Consequently, the deviations from these targets are calculated as the absolute sum of linear differences above and below the target value as there is not a tolerance nor, obviously, upper or lower bounds. Similarly, there is not a residual district. In all three cases, the target size of districts vary for each Census Tract Part, depending upon the supplied number of districts. As is the case for PFOM and dwelling based approaches, however, the average deviations for perimetre or surface area characteristics must be less than 10% of the target size. But, because the model works with 'units of one' in the case of blocks, a more reasonable constraint/tolerance was chosen to be that the average deviation in this case be less than the supplied target number of districts. This permits, on average, a deviation of one block per district (which typically contains between 15 and 20 blocks).

Districtings with average deviations that are less than or equal to the target average deviation pass the evaluations and those that do not, fail.

It should be noted that the results shown in the following sections reflect the final runs for each module. (That is, some tests had to be re-run after amendments/corrections were made to the autodistricting programs.)

5.3 TEST PATTERN DATA SETS

A series of test data sets of controlled variation in street layouts and dwelling distributions were used to ensure the proper functioning of the implemented districting models. The test pattern sites developed for this type of testing are depicted in Figure 5.1.

The characteristics of each of the "test pattern" sites are summarized in Table 5.1. Three of the 'test Census Tract Parts' have uniformly dense street network layouts (CT #901, CT #902, and CT #998) and one of these (CT #901) has a uniform distribution of dwellings. CT #991 has the majority of its blocks skewed into one corner of the test area. CT #993 has a set of blocks of increasing size and decreasing dwelling density that fan out in a circular/ring-like fashion from the centre. Finally, CT #992 has a variable distribution of clusters of blocks that are akin to subdivision developments on the periphery of most large Canadian cities. Together, these test sites are representative of the variety of settlement distribution, pattern and density that are found in most Canadian cities. (Note: the distribution of the block centroids, and not the shape of the blocks is the most important variant.)

To validate the functioning of the programs, each of the nine methods described in Chapter 2 was tested, for each of the 5 optional characteristics or 'possible objective functions', (i.e., block counts, perimetre, surface area, dwelling counts, and elapsed time), for the aggregation approach and once (since the districts form are identical for all characteristics) for the disaggregation approach on the test data set for CT #901. This produced a total of 54 districtings of the highly regular/controlled CT #901.

The nature of the testing for verification on the 6 "test patterns" and the results of the 54 tests on the first test pattern, CT #901, are described in the next section. The results for the block, perimetre, area and elapsed time based districting are also assessed (i.e., given a pass or fail) for CT #901 and expressed in terms of the dwelling count distributions. The final results of the dwelling count aggregation and the partial figure of merit of aggregation tests for the remaining test pattern Census Tracts are also provided.

Figure 5.1 Model Test Census Tracts

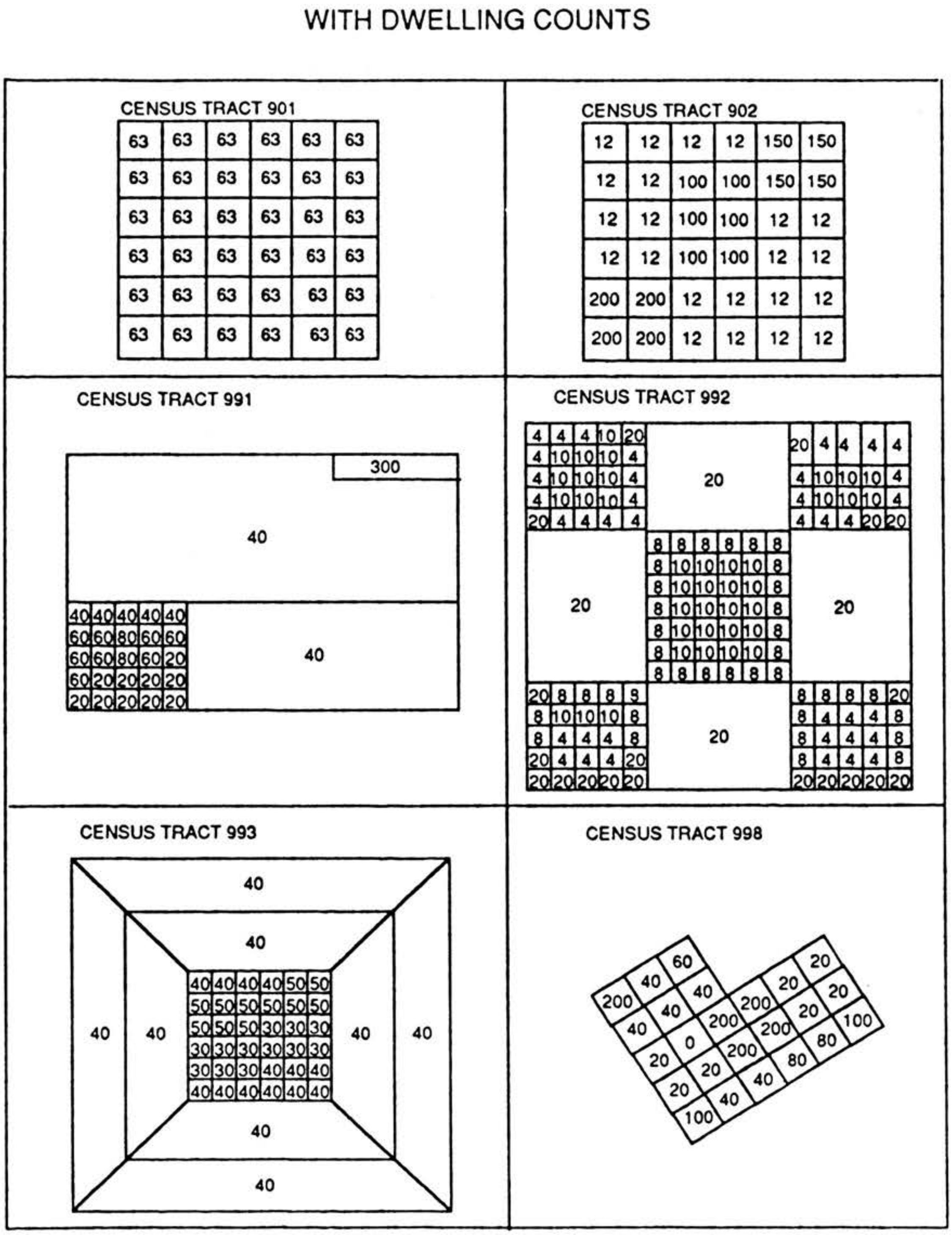


TABLE 5.1 CHARACTERISTICS OF MODEL SITES

CHARACTERISTICS	SITE NUMBER					
	CT #901	CT #902	CT #991	CT #992	CT #993	CT #998
SETTLEMENT DENSITIES	Uniform	Varied	Varied	Relatively Uniform	Relatively Uniform	Varied
NUMBER OF DWELLINGS	2268	2264	1640	1326	1750	1800
PERIMETRE (km)	540	540	3633	13073	8511	3162
STREET PATTERN	Uniform	Uniform	Skewed	Varied	Circular	Uniform
LINGUISTIC DIVERSITY	English and French	English and French	English and French	English and French	English and French	English and French
SURFACE AREA (sq. km.)	Small 2.2	Small 2.2	Very Large 457.3	Very Large 497.8	Very Large 662.6	Medium 115.7
NUMBER OF BLOCKS	36	36	28	141	44	24
AVERAGE BLOCK DWELLING COUNT	63	63	59	9	40	75
DWELLING DENSITY (sq km)	1008	1006	4	3	3	16
DWELLINGLESS BLOCKS	0	0	0	0	0	1
BLOCK-FACE CENTROIDS	36	36	28	141	44	24
TARGET NUMBER OF DISTRICTS*	7	7	5	4	5	5
EXPECTED** RESIDUAL COUNT	18	14	140	201	250	300

* The target number of districts is determined by dividing the number of dwellings by 375 and rounding upwards to the next whole/integer number.

** The expected residual count is determined by subtracting the product of the target number of dwellings (375) and the target number of districts minus one from the number of dwellings.

5.4 MODEL VALIDATION TESTING

5.4.1 Results Of Model Validation Testing

The results of the 54 tests (9 for disaggregation and 45 for aggregation) on CT #901 are summarized in Table 5.2. All districtings based on aggregating blocks, perimetres and surface areas passed (if contiguity is not considered) the evaluations based both on an assessment of the geographical distribution of the individual characteristic (i.e., blocks, perimetres, surface areas and elapsed time) and of the dwellings. All districtings based on aggregating dwellings also passed (excluding discontiguities). However, the fifth (rings with sectors) and sixth (bidirectional grid) methods failed to produce acceptable districtings using the PFOM-based aggregation approach.

Conversely, only the sixth (bidirectional grid), seventh (regular seeds) and ninth (unidirectional grid) methods produced an acceptable districting based on disaggregation of dwelling counts.

It is clear, therefore, from Table 5.2 that disaggregation methods which essentially "force-fit" theoretical models to the surface (i.e., assume a continuous or locally balanced distribution of the selected variable), performed much more poorly than aggregation methods which combined discrete elements of the actual surface in a more flexible fashion.

Disaggregations -- assessed in terms of the distribution of dwellings and ignoring discontiguities -- succeeded in only 3 of 9 cases (33%). Aggregations -- assessed on the same basis -- succeeded in 43 of 45 cases (96%). As a result, further (feasibility) testing focused only on aggregative methods.

Assessed on the basis of the individual characteristics other than dwelling counts, aggregation approaches succeeded in 34 of 36 cases (94%) as shown in Table 5.2. Each districting (one for each of the nine methods and for each characteristic) is evaluated on the basis of the characteristic and on the basis of the dwelling count distribution. A 'P' is placed in the table if the districting meets the criteria and an 'F' if it does not. The table also records the total number of discontinuous parts (or subdistricts) for each districting.

On average, 4 of the 9 different districtings methods generated results without discontiguities for each aggregation characteristic. On average, 3 of the 9 methods had only one discontiguity. Thus, if 3 of 9 methods are used, it is likely (i.e., greater than 50%) that there will be at least 1 solution is without any discontiguities. It is also highly probable (at least 84% in this case) that there will be 1 solution with at most 1 discontiguity.

Table 5.2 Results of Assignment Methods on CT #901

Approach	Method Number #1	#2	#3	#4	#5	#6	#7	#8	#9
DISAGGREGATION									
Dwelling-Based	F	F	F	F	F	P	P	F	P
Discontiguities (Parts)	0	0	12	0	0	0	1	0	0
AGGREGATION									
Block Counts									
Blocks	P	P	P	P	P	P	P	P	P
Dwellings	P	P	P	P	P	P	P	P	P
Discontiguities (Parts)	2	0	18	1	1	0	1	3	0
Perimeters									
Length	P	P	P	P	P	P	P	P	P
Dwellings	P	P	P	P	P	P	P	P	P
Discontiguities (Parts)	0	0	18	1	2	0	1	2	0
Surface Area									
Area	P	P	P	P	P	P	P	P	P
Dwellings	P	P	P	P	P	P	P	P	P
Discontiguities (Parts)	0	0	18	1	1	0	1	8	0
Dwelling Counts									
Dwellings	P	P	P	P	P	P	P	P	P
Discontiguities (Parts)	0	0	18	1	1	0	1	5	0
Partial FOM									
PFOM	P	P	P	P	F	F	P	P	P
Dwellings	P	P	P	P	F	F	P	P	P
Discontiguities (Parts)	0	0	18	1	0	0	1	3	0

"P" = PASS, "F" = FAIL, "(Parts)" = Number of Discontiguities

Number	Method	Number	Method
1	Extrema-based Seeds	6	Bi-directional Grids
2	Sectors	7	Regular Seeds
3	Rings	8	Random Seeds
4	Sectors with Rings	9	Uni-directional Grids
5	Rings with Sectors		

The results of applying each of the nine aggregation methods to the remaining 'test Census Tracts' using the dwelling count approach are summarized in Table 5.3. The PFOM approach was also tested in one additional case, CT#902, to see if it would perform relatively better on a less uniform pattern of dwelling distributions, since it was the only approach, aside from the DISAGGREGATION approach, that did not enjoy much success in early testing on the uniform pattern of CT #901.

1. PFOM-Based Aggregations

Using PFOM-based aggregation, only Method #1 (extrema-based seeds) produced a successful districting without discontiguities. Two methods produced acceptable districtings with 3 and 4 discontiguities respectively and the remaining six methods failed to produce an acceptable result.

Since the focus for these tests was on the proper functioning of the programs, and since the programs being tested were parameterized to use any one of the five different characteristics, further validation testing of the PFOM-based aggregation was limited to CT #902.

2. Dwelling Count-Based Aggregations

Successful districtings without discontiguities were produced by at least one method for all CTPs except CT #992. Two districtings for CT #992 have exactly 1 discontiguity and three have 2 discontiguities. Methods #1, #5 and #6 were able to successfully district at least 2 test pattern "Census Tracts" without discontiguities.

While the number of discontiguities proved to be generally small, (i.e., 1-3 discontiguities in 25 of 35 or 73% of the cases), only 10 of the 45 or 22% of the test cases have solutions with no discontiguities.

Table 5.3 Results of Assignment Methods on CT #902, CT #991, CT #992, CT #993 and CT #998.

Census Tract	Method Number								
	#1	#2	#3	#4	#5	#6	#7	#8	#9
Census Tract #902									
Dwellings	P	P	F	F	F	P	P	P	F
Discontiguities (Parts)	0	4	16	2	3	0	2	6	2
PFOM	P	P	F	F	F	F	F	P	F
Discontiguities (Parts)	0	4	15	1	0	2	5	3	2
Census Tract #991									
Dwellings	P	P	P	F	F	P	P	P	P
Discontiguities (Parts)	0	4	2	2	3	0	2	1	0
Census Tract #992									
Dwellings	P	P	P	P	P	P	P	P	P
Discontiguities (Parts)	4	1	7	2	3	2	2	4	1
Census Tract #993									
Dwellings	P	P	P	P	P	P	P	P	P
Discontiguities (Parts)	1	2	8	0	0	2	7	0	1
Census Tract #998									
Dwellings	P	P	P	P	P	F	P	F	F
Discontiguities (Parts)	1	0	6	1	0	1	1	2	2

"P" = PASS, "F" = FAIL, "(Parts)" = Number of Discontiguities

Number	Method	Number	Method
1	Extrema-based Seeds	6	Bi-directional Grids
2	Sectors	7	Regular Seeds
3	Rings	8	Random Seeds
4	Sectors with Rings	9	Uni-directional Grids
5	Rings with Sectors		

5.4.2 Conclusions From Model Validation Testing

The validation tests confirmed that the autodistricting algorithms/programs function as expected. They also indicated that all methods, with the exception of the method based on concentric rings, generated useful districtings with relatively few discontiguities.

Most of the districtings were automatically assessed as acceptable (i.e., 'PASS') which halts the further search for an acceptable districting in a production environment. However, for the validation testing, in order to assess the method selection procedure, it was necessary to continue the process until all methods were applied to each site. The result from each method was ranked in terms of acceptability to determine the three best methods for the given site.

The expected occurrence of small numbers of discontiguities initially justified the earlier decision to develop and test the 'RE-ASSIGN process' described in Chapter 4. The RE-ASSIGN process eliminates the occurrence of discontiguous district parts. However, the RE-ASSIGN process, at that time, led to oversized districts (and subsequently, when changes were made to the decision rules, led to extra districts). Many of the discontiguities involved blocks containing no dwellings and which were eventually dealt with by procedures for 'bonding' individual blocks prior to districting and groups of blocks after districting using the ANNEX Module.

These early indications of the limitations of a 1-pass ASSIGN approach and the relative costliness of the RE-ASSIGN process eventually led to the design of a 2nd pass ANNEX process to regroup the parts/subdistricts of a proposed districting to form contiguous districts. The ANNEX process was described in Chapter 4.

5.5 FEASIBILITY CASE STUDY TESTING

The ability to compare districtings produced by computer-assisted means with those produced by traditional manual methods greatly increases the relevance of the results.

A second set of tests, called feasibility case studies, was undertaken to assess the performance of the model relative to the actual and the manually simulated results for the 1986 Census, as produced by traditional manual methods. This section indicates how the test sites were selected, describes their nature and documents the results of the testing.

5.5.1 The Site Selection Process (circa October 1984)

The selection of case study sites to demonstrate feasibility was made prior to the 1986 Census and this added a number of pragmatic considerations to the general criteria employed in selecting appropriate sites.

Given large volumes of data (cartographic and detailed statistical data) needed for each test site, and given the limited time and financial resources thought to be available at that time, the selection of suitable test sites required great care. A number of factors were considered:

1. timely availability of comparative 'manually generated' districts for the 1986 Census;
2. timely availability of up-to-date (and upgraded) cartographic files for the 1986 Census;
3. sufficient linguistic variety to test that component of the model;
4. variety in the settlement densities;

5. diversity in the street network patterns; and
6. stability in regional extent (i.e., Census Tract limits) over successive censuses to permit (eventual) longitudinal analysis of the model's performance.

Comparing production schedules for manual collection unit districting of the 1986 Census and for the updating of cartographic base files, with the original planned timeframe for testing the model, eliminated the vast majority of a possible 43 Census Tracted centres due to data unavailability.

Of the set of regions with completed upgrades of the cartographic base files and the early availability of comparative manual districtings, only one, Laval, exhibited sufficient linguistic variety to be able to test that component of the model. Laval is one of the municipalities comprising the Census Metropolitan Area of Montreal in the Province of Quebec.

The Census Subdivision (CSD) of Laval is subdivided by three Federal Electoral District (FED) boundaries and contains a mix of urban and rural landscapes. The Laval CSD is an island and thus easily separated from surrounding municipalities by channels or tributaries of the St. Lawrence River.

The municipality of Laval comprised 61 Census Tracts in 1986. A set of four (4) CT selected sites were used:

1. to test the model; and
2. to fine tune the model.

Within that set of Census Tracts and after receiving input from the staff responsible for the maintenance of the Census Tract Program, it was decided that CT #701, CT #702, CT #707 and CT #715 would be selected to provide the desired diversity in settlement density and in complexity of street pattern layout for the first phase of testing.

The landscape for each of these four (4) Census Tracts is depicted in photographs shown in Illustrations 5.1, 5.2, 5.3 and 5.4 respectively. It is clear from even these few examples that there is a broad diversity in settlement type and pattern in the selected Census Tracts. The variability is further documented in Table 5.4.

Illustration 5.1: Images of Laval Census Tract #701 showing agricultural land use.

Illustration 5.2: Images of Laval Census Tract #702 showing a typical urban residential pattern.

Illustration 5.3: Images of Laval Census Tract #707 showing higher density urban settlement.

The character of each of these study sites is summarized below in Table 5.4.

TABLE 5.4 CASE STUDY SITE CHARACTERISTICS

CHARACTERISTIC	CENSUS TRACT			
	CT #701	CT #702	CT #707	CT #715
SETTLEMENT DENSITIES	Variable Settled - Vacant	Variable Settled - Vacant	Variable Dwelling Densities	Variable Urban - Rural
NUMBER OF DWELLINGS	1966	2092	1867	2108
PERIMETER (km)	1263	645	574	1218
STREET PATTERN	Concentric	Skewed	Uniform	Variable
LINGUISTIC DIVERSITY	Mainly French Speaking	Mainly French Speaking	Bilingual	Mainly French Speaking
GEOGRAPHIC AREA (sq. km.)	Medium 9.5	Medium 8.2	Small 1.2	Large 12.9
NUMBER OF BLOCKS	87	44	61	57
AVERAGE BLOCK DWELLING COUNT	23	48	31	37
DWELLING DENSITY (sq km)	207	254	1650	163
DWELLINGLESS BLOCKS	41	14	7	14
BLOCK-FACE CENTROIDS	154	103	164	240
DWELLING COUNT WITH FILTERING OF SUB-BLOCKFACE UNITS	1166	1692	1867	1708
TARGET NUMBER OF DISTRICTS*	4	5	5	5
EXPECTED RESIDUAL COUNT**	41	192	367	208

* The target number of districts is determined by dividing the dwelling counts (after filtering of sub-blockface units) by the target number of dwellings (375).

** The expected residual is determined by subtracting the product of the target number of dwellings and the target number of districts minus one from the dwelling counts (after filtering of sub-blockface units).

A plot of the Census Tracts contained within the Census Subdivision of Laval showing the relative location of the selected study sites is provided as Figure 5.2. Figures 5.3 through 5.6 are computer produced plots of each of the case study site locations showing street patterns and block dwelling counts.

FIGURE 5.2 LAVAL CENSUS TRACTS (PARTS)

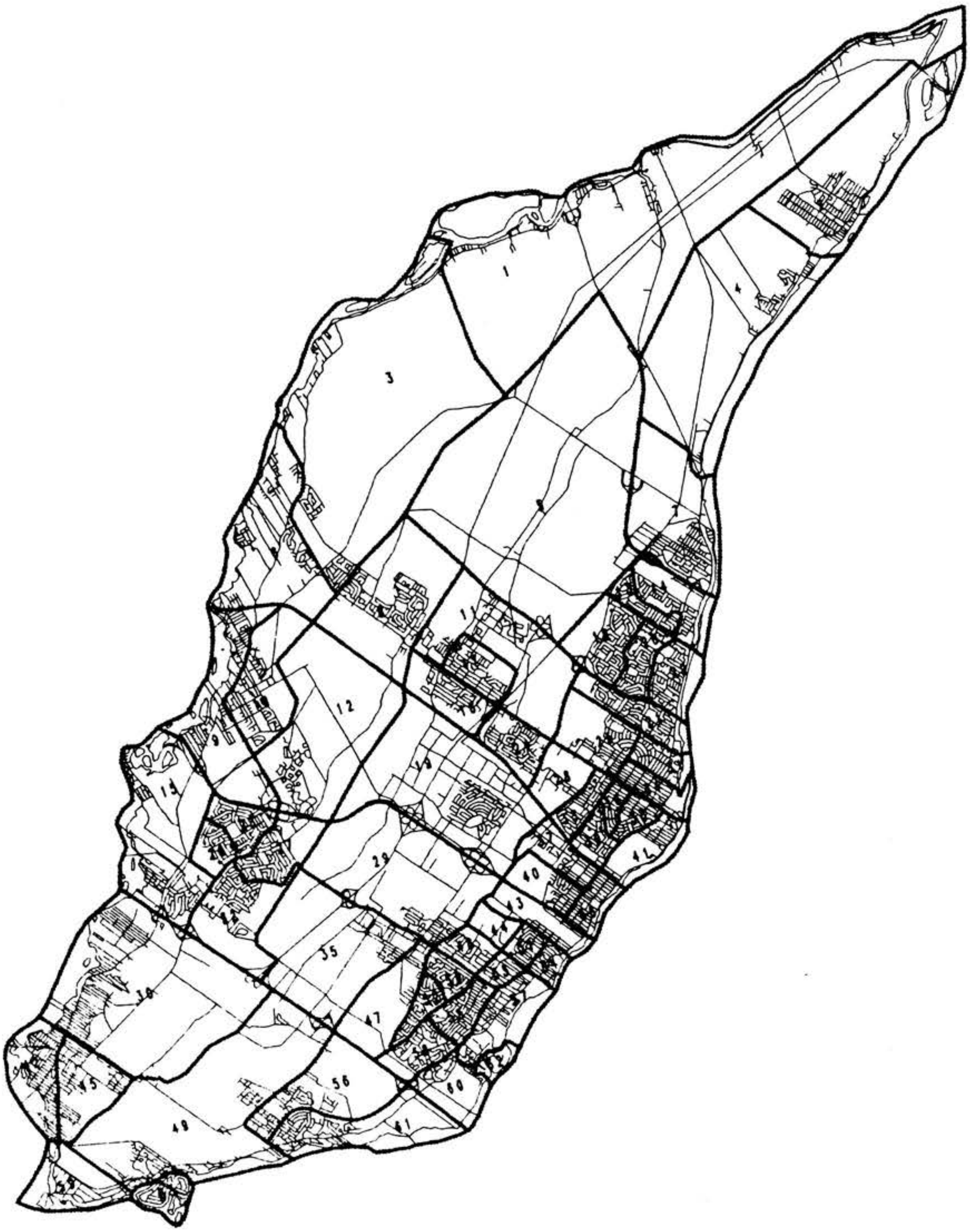


FIGURE 5.3 LAVAL CENSUS TRACT #701

FIGURE 5.4 LAVAL CENSUS TRACT #702

FIGURE 5.5 LAVAL CENSUS TRACT #707

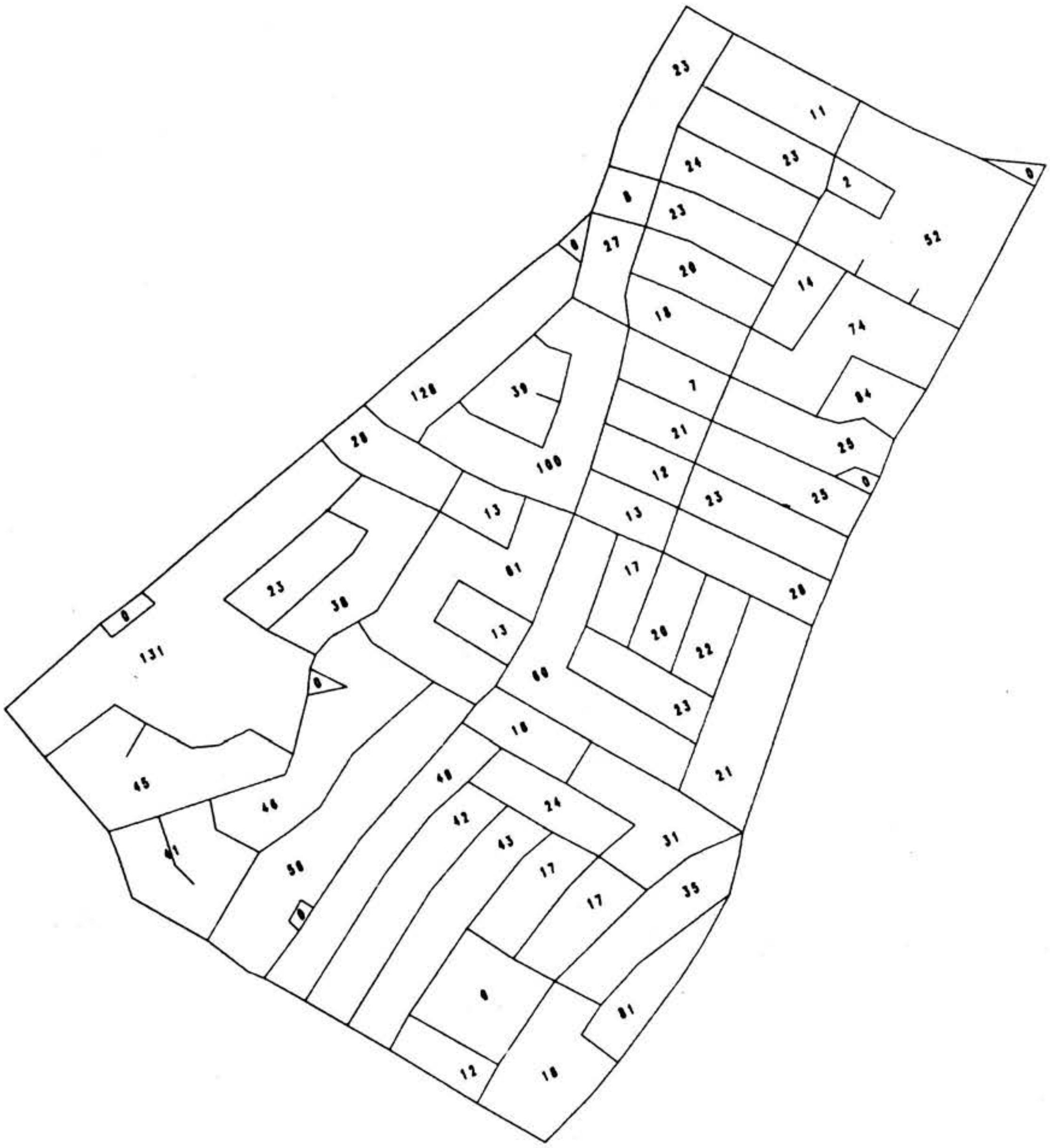
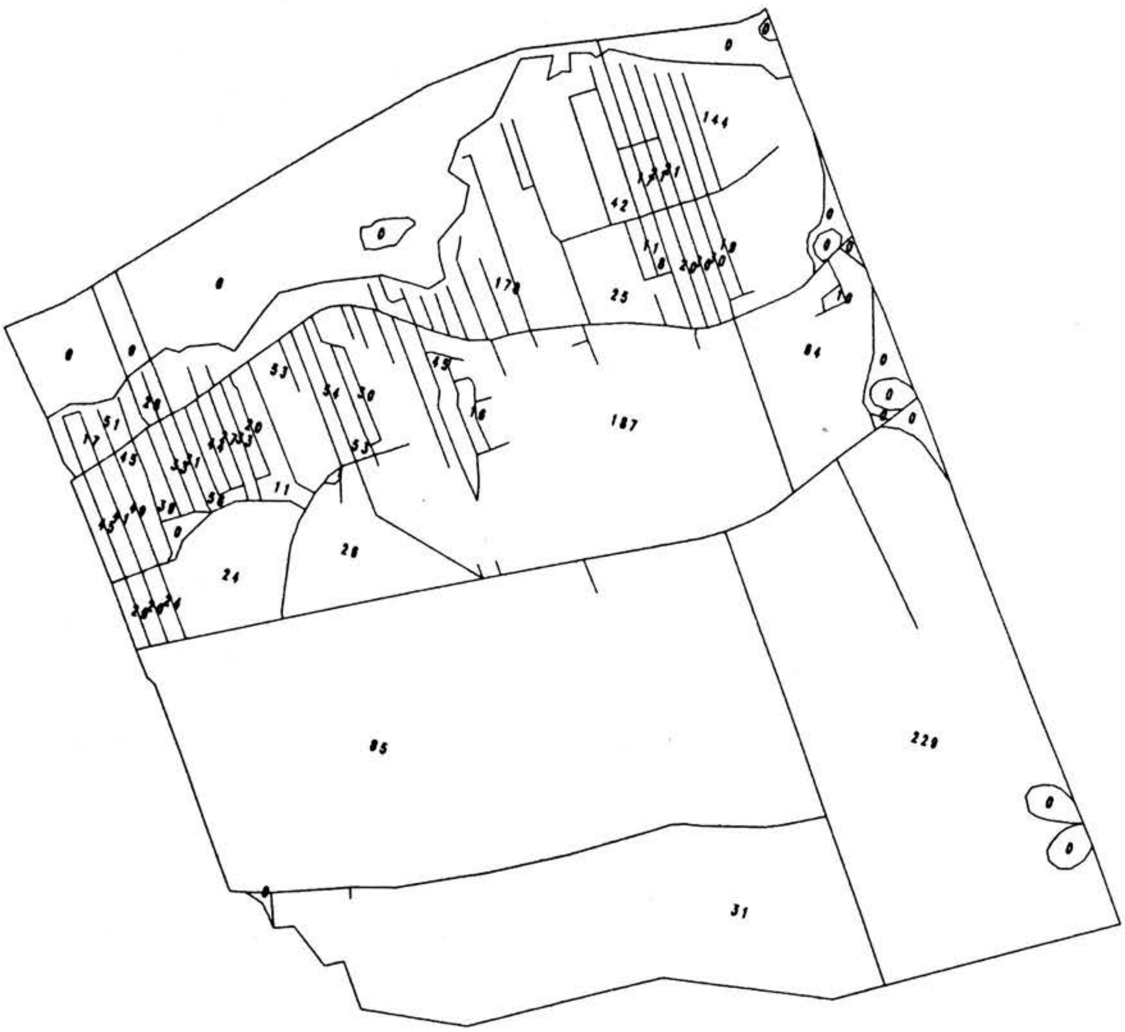


FIGURE 5.6 LAVAL CENSUS TRACT #715



5.6 VIABILITY CASE STUDY RESULTS

Again, to permit the subsequent assessment of the method selection procedure, all methods were applied to each of the four CTPs in the sample. The results of applying each of the nine AGGREGATION ASSIGN methods to the four (4) actual Census Tract Parts from Laval using the dwelling based approach are summarized in Table 5.5 (i.e., results from the ANNEX and RE-ASSIGN processes are not yet presented).

Table 5.5 Results of AGGREGATION ASSIGN Methods on CT #701, CT #702, CT #707, and CT #715 (before the ANNEX and RE-ASSIGN processes)

Census Tract	Method Number								
	#1	#2	#3	#4	#5	#6	#7	#8	#9
Census Tract #701									
Dwellings	P	P	P	P	P	P	F	F	P
Discontiguities (Parts)	4	3	2	5	5	2	8	3	9
Census Tract #702									
Dwellings	P	P	P	F	P	P	P	F	P
Discontiguities (Parts)	2	0	5	0	4	1	6	3	2
Census Tract #707									
Dwellings	P	F	F	P	P	P	P	P	F
Discontiguities (Parts)	2	0	11	4	2	2	2	7	3
Census Tract #715									
Dwellings	P	P	P	P	F	F	P	P	P
Discontiguities (Parts)	3	3	6	5	7	2	8	3	3

"P" = PASS, "F" = FAIL

Number	Method	Number	Method
1	Extrema-based Seeds	6	Bi-directional Grids
2	Sectors	7	Regular Seeds
3	Rings	8	Random Seeds
4	Sectors with Rings	9	Uni-directional Grids
5	Rings with Sectors		

Excluding consideration of discontiguities, only Method #1 (extrema based seeds) produced a "PASS", in terms of dwelling counts, on all four Census Tracts. Conversely, all methods generated a "PASS" solution in at least 3 of 4 cases (75%) except #8, random seeds (which had a "PASS" solution in 50% of the cases).

From an operational perspective, the number of discontiguities reduced the perceived acceptability of the ASSIGN process component of the districting model. While the RE-ASSIGN process was able to eliminate all discontiguities, it was judged to be cost prohibitive to utilize the RE-ASSIGN capability for each use of an ASSIGN process. The ANNEX process, which was described in the previous chapter, was therefore designed/specified by the author and implemented by the autodistricting team (during the viability case study stage) as a very cost effective addition to the districting tool kit. At this juncture, the focus was placed on optimizing the districting process at the level of the individual collection unit and not at the level of the actual workload assignment. As mentioned earlier, this represents a slightly higher standard than actual manual districting. (While the creation of actual "workloads" is an obvious extended application of the autodistricting model, grouping of districts to form "workloads" is currently done manually.)

While quality considerations are an important element of viability (and while the quality of the results for the viability testing were acceptable, especially after the incorporation of the ANNEX process), cost and throughput are also important. In the manual approach, elapsed time for districting has a direct bearing on both cost and throughput. Therefore, an assessment of manual effort typically required for the Laval case study Census Tracts was undertaken.

5.6.1 Manual Districting Results For Case Study Areas

Although districting results from the 1986 Census were readily available, detailed cost estimates were not. Therefore, each of the four (4) case study census tracts was also manually districted by three (3) different members of the districting staff to establish a benchmark for cost comparisons and to gain an appreciation of the range of "acceptable districtings". The time taken for manual districting is shown in Table 5.6.

Because two districtings for CT #707 were performed by different staff than for the remaining Census Tracts, the results are excluded from this comparison. The resulting districtings are shown in Figure 5.7. In all cases, the manual districting passed the dwelling PASS/FAIL evaluation function. In each case, however, the districting staff chose to split blocks to achieve the desired results. This would seem to imply that the selected sample represents a set of relatively difficult cases since written procedures discourage the splitting of blocks. Autodistricting incorporates, as necessary, block splits carried forward from the previous Census and allows for the introduction of new splits via interactive tools in the ADJUST module.

TABLE 5.6 DISTRICTING TIMINGS BY CENSUS TRACT BY EXPERTS

	CENSUS TRACT NUMBER		
STAFF MEMBER	CT #701	CT #702	CT #715
94	3 hours	2 hours	2 hours
95	2 hours	2 hours	3 hours
96	2 hours	0.5 hours	1 hour
Mean	2.3 hours	1.5 hours	2 hours
Mode	2 hours	2 hours	2 hours
Variance	0.2	0.5	0.7
Standard Deviation	0.5 hours	0.7 hours	0.8 hours

FIGURE 5.7 MANUAL DISTRICTINGS OF LAVAL CENSUS TRACTS

CENSUS TRACT #701

EA #	# Dwellings
1	385
2	391
3	393
4	402
5	395

2 block splits
PASS

EA #	# Dwellings
1	386
2	393
3	397
4	398
5	392

2 block splits
PASS

EA #	# Dwellings
1	221 (residual)
2	370
3	389
4	379
5	380
6	227

3 block splits
PASS

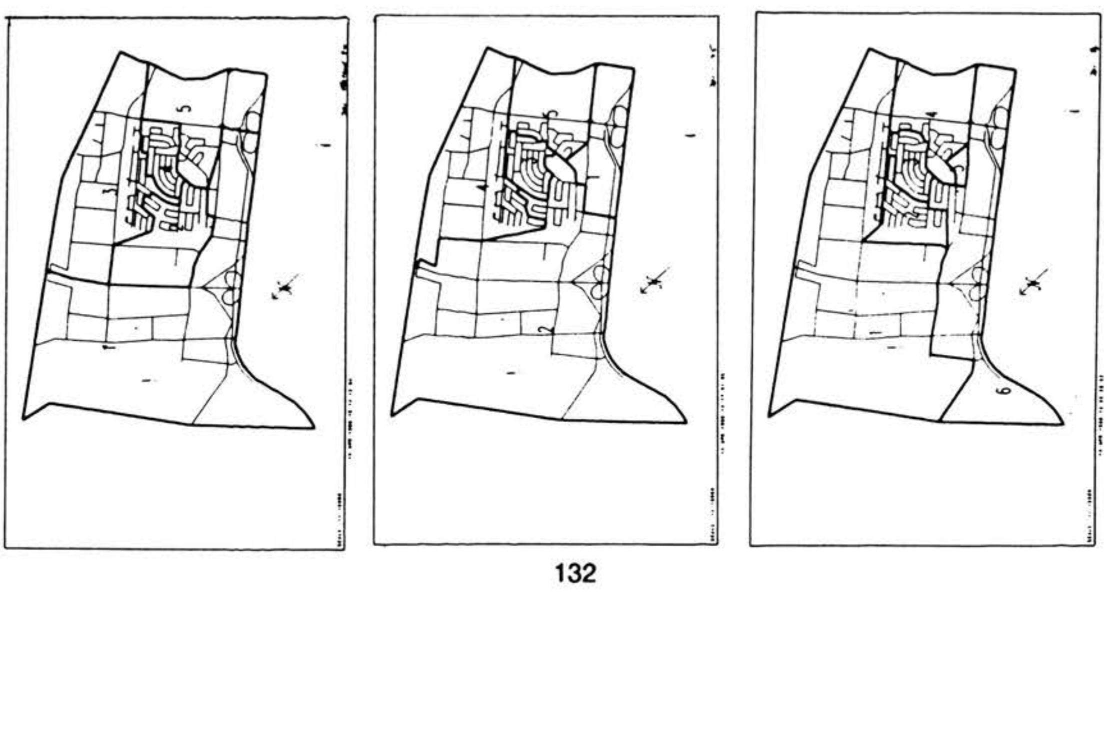
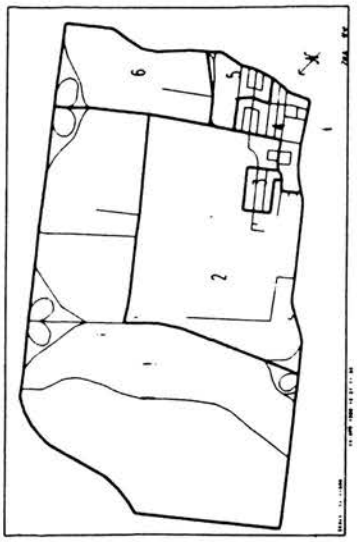


FIGURE 5.7 (Continued)

CENSUS TRACT #702

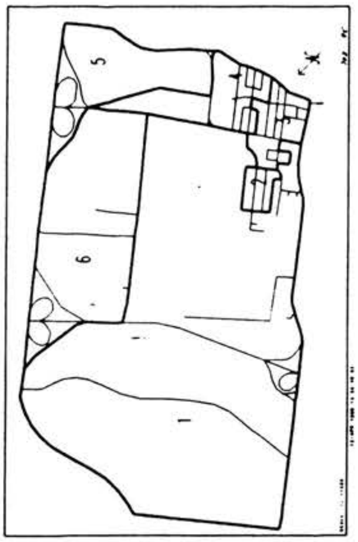
EA #	# Dwellings
1	262 (residual)
2	384
3	378
4	397
5	369
6	302

PASS



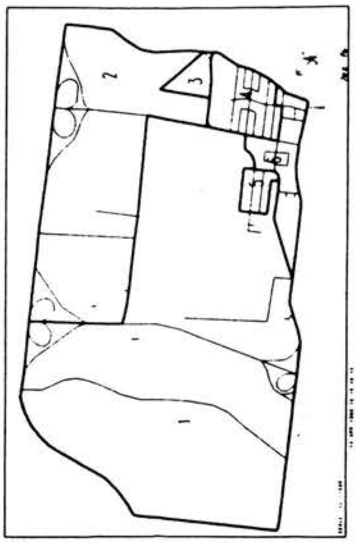
EA #	# Dwellings
1	383
2	378
3	366
4	340
5	400
6	225 (residual)

PASS



EA #	# Dwellings
1	268 (residual)
2	327
3	400
4	364
5	378
6	355

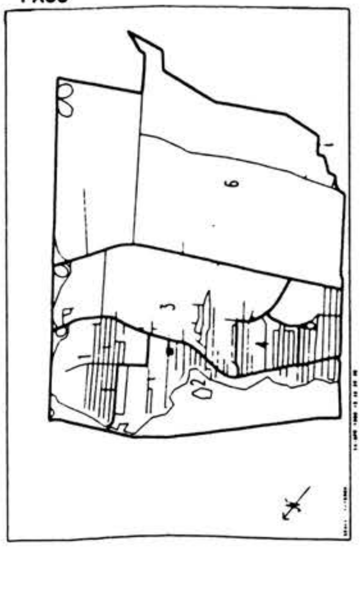
PASS



CENSUS TRACT #715

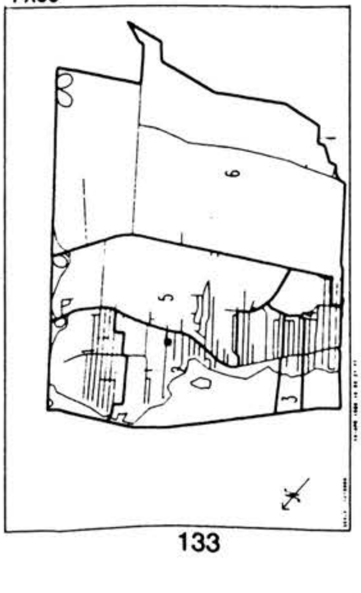
EA #	# Dwellings
1	359
2	303 (residual)
3	349
4	381
5	371
6	345

2 block splits
PASS



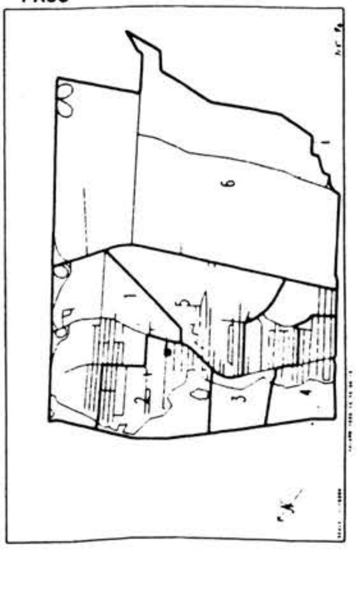
EA #	# Dwellings
1	376
2	392
3	193 (residual)
4	383
5	379
6	385

6 block splits
PASS



EA #	# Dwellings
1	371
2	280 (residual)
3	373
4	370
5	369
6	345

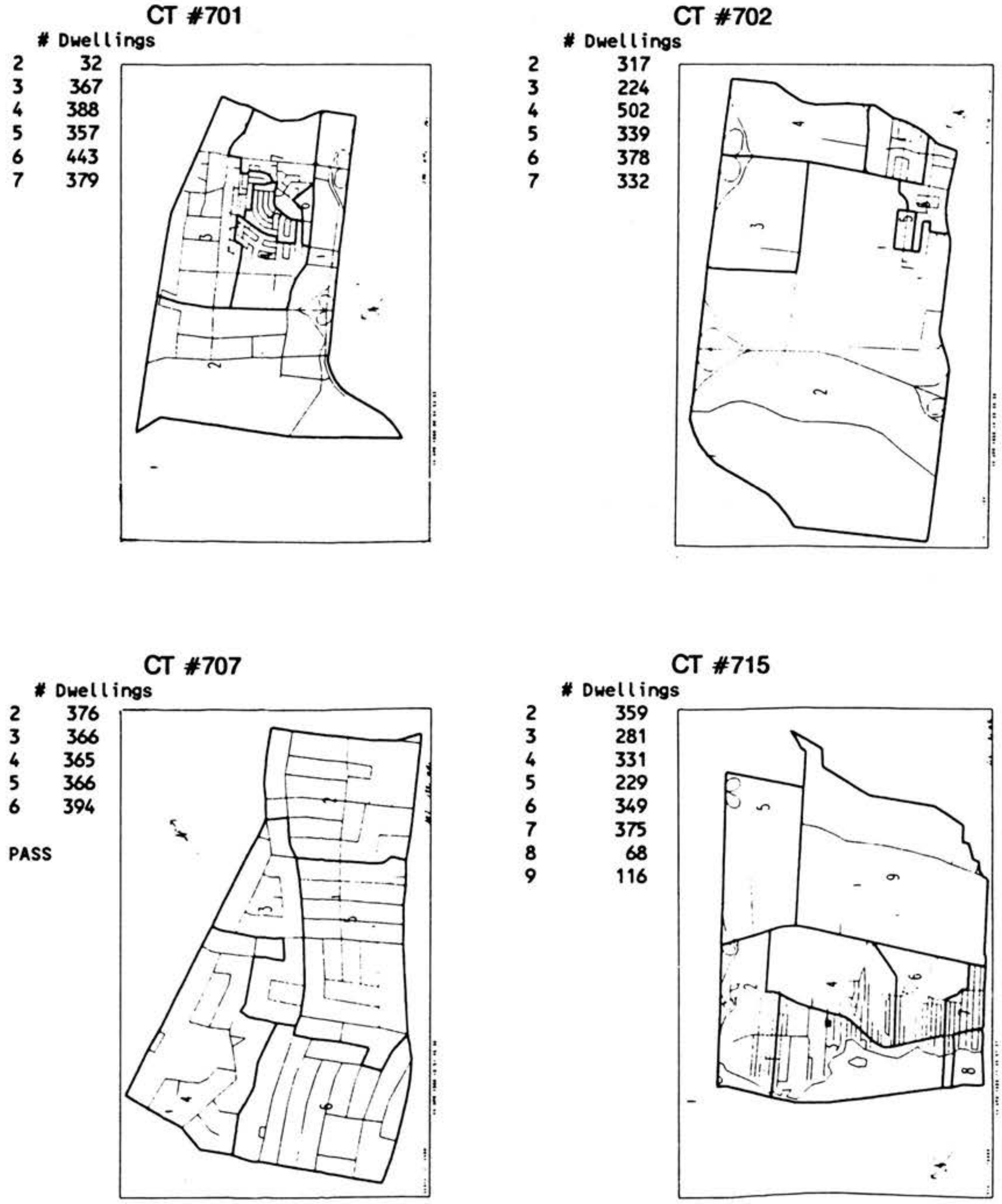
7 block splits
PASS



5.6.2 Previous Districting Results For Case Study Areas

Both the manual and the automated system approaches first assess the appropriateness of using the same districting employed during the previous (1981) Census, (although this was obviously not possible for the test pattern districts described in the previous section). The previous districtings are shown in Figure 5.8.

Figure 5.8 Previous (1981) Census Districtings for Test Census Tracts



The districtings from the previous (1981) Census for CT #702 and CT #707 "passed" the dwelling based evaluation function and were still acceptable. The district limits for those CTs would be, therefore, re-employed and redistricting was required only for CT #701 and CT #715.

5.6.3 Autodistricting Results For Case Study Areas

The automated selection of ASSIGN processes based on a special coefficient of variation value of 0.3932 and on Table 4.5 resulted in the following techniques being selected for CT #701:

Table 5.7 Autodistricting Results for CT #701

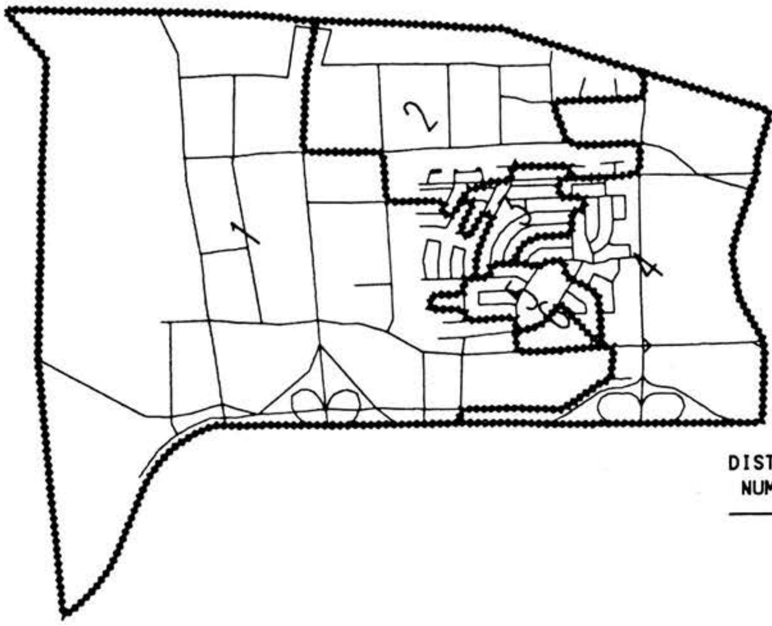
Steps	Methods	Descriptions	Status	Deviations
Assignments				
First	6A*	Bi-directional Grid	PASS	12.5
Second	1A*	Extrema-Based Seeds	FAIL	122.0**
Third	2A*	Sectors	PASS	0.0

* the 'A' indicates that the results from the ASSIGN process have also passed through the ANNEX process.
** Subdistricts are treated as districts.

In a production environment, the districting process would stop as soon as an acceptable solution was generated. For method selection testing purposes, all methods (and, later, the results of applying the ANNEX process to the result) were generated. The initial results of this testing procedure are provided in Table 5.5 and the final results for CT #701 (i.e., after using the ANNEX process) are provided in Table 5.7 and are depicted in Figure 5.9. Thus, for example, for CT #701, Method #6 had 2 discontiguities while Method #6 after the ANNEX process had none.

Figure 5.9 Assignment Methods for CT #701

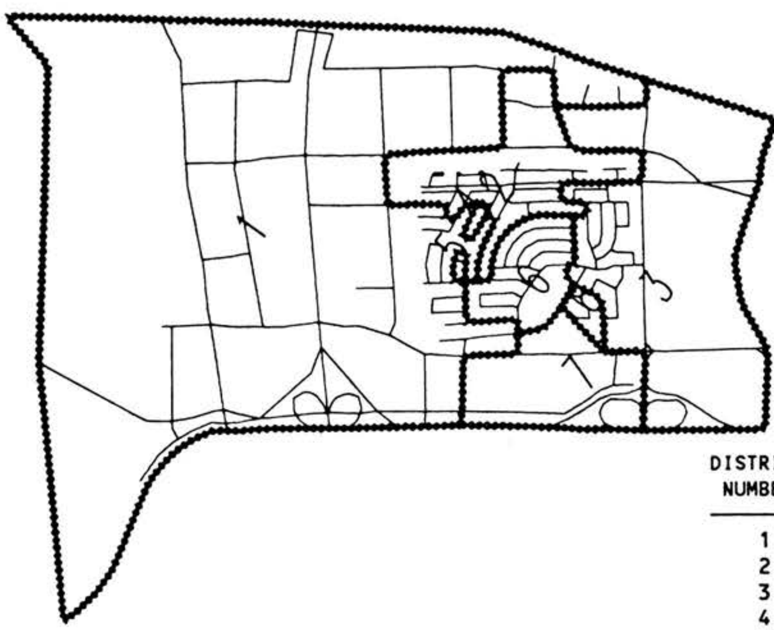
First Assignment Method for CT #701 - Bi-directional Grids



TARGET = 6

DISTRICT NUMBER	DWELLING COUNT	PFOM
1	386	2046.968
2	204	1053.574
3	317	1361.213
4	371	1779.922
5	308	1286.242
6	380	1474.633

Second Assignment Method for CT #701 - Extrema-Based Seeds

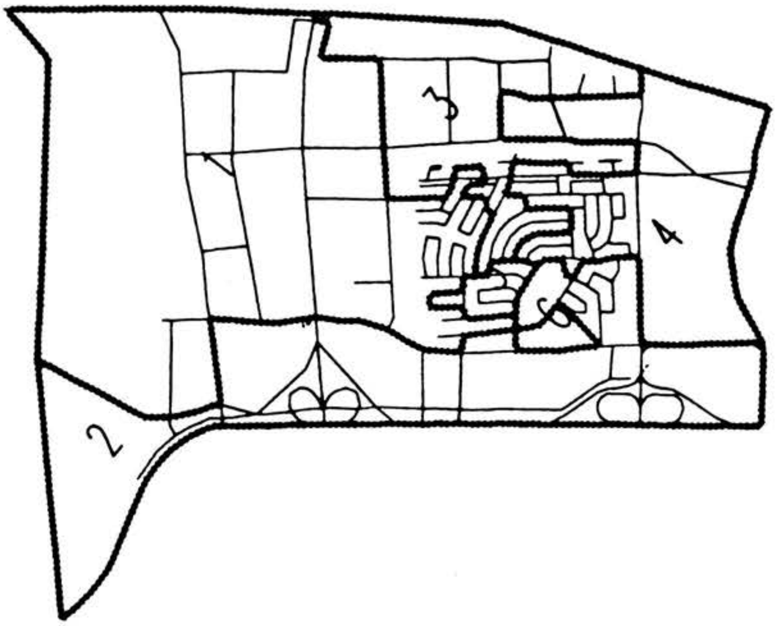


TARGET = 6

DISTRICT NUMBER	DWELLING COUNT	PFOM
1	388	2142.451
2	388	1731.583
3	303	1461.670
4	22	94.454
5	16	69.712
6	292	1256.093
7	385	1558.843
8	172	687.745

Figure 5.9 (Continued)

Third Assignment Method for CT #701 - Sectors



TARGET = 6

DISTRICT NUMBER	DWELLING COUNT
1	378
2	362
3	367
4	351
5	128
6	380

PASS

The method selection for ASSIGN processes resulted in the following techniques being employed for CT #715 based on a special coefficient of variation value of 0.3433 and Table 4.5:

Table 5.8 Autodistricting Results for CT #715

Steps	Methods	Descriptions	Status	Deviations
Assignments				
First	6A	Bi-directional Grids	FAIL	55.7
Second	1A	Regular Seeds	FAIL	45.5
Third	2A	Sectors	FAIL	20.7*
Re-Assignments				
First	2A	Reassign on Sectors	FAIL	454.0
Second	1A	Reassign on Extrema-Based Seeds	FAIL	54.6
Adjustment				
First	2A	Exemption on Node Joins for Sectors	PASS	7.0**

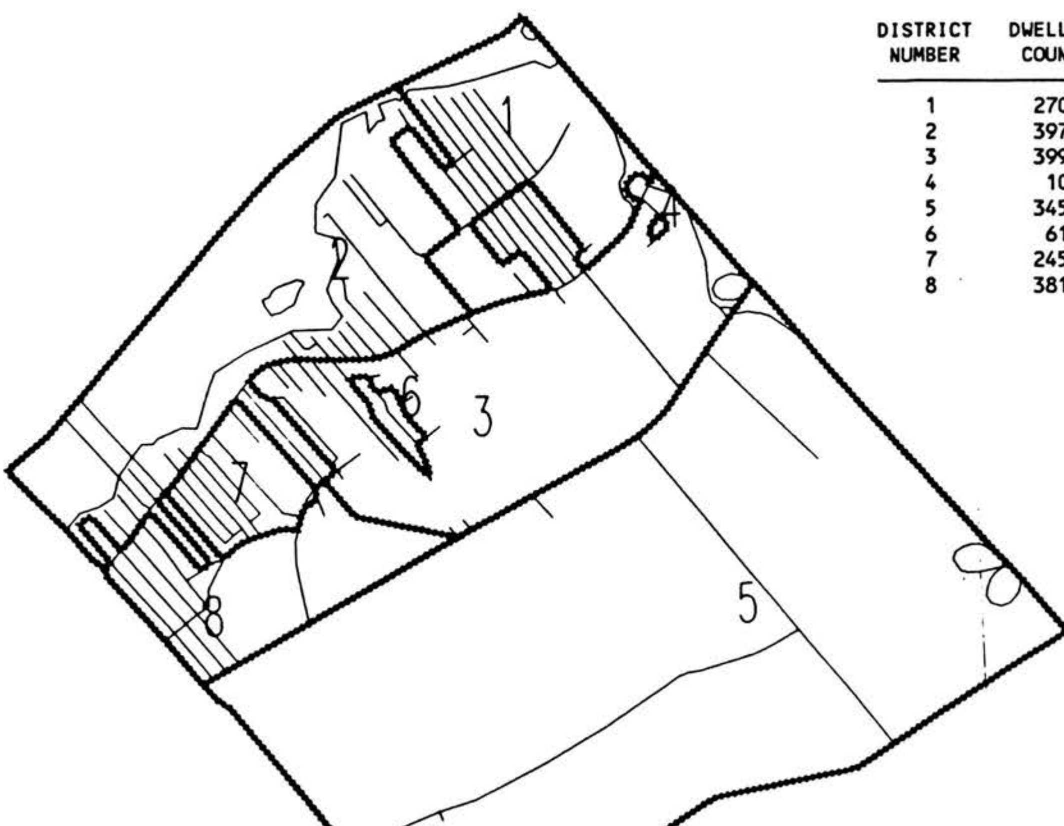
* The FAIL status was assigned based on the number of districts and not on the fact that the average deviation was over 37.5.

** The acceptance of two subdistricts being joined by a common node reduced both the number of districts and the average deviation.

In this case, the complexity of this Census Tract Part meant that all steps in the districting process were unsuccessful. Finally, the ADJUST process granted an exception (i.e., no changes were needed other than to consider two "connected" subdistricts to be a single unit) to allow an earlier result (ANNEX of Method #2 - Sectors) to be accepted even though the two subdistricts (#7 and #8 in Figure 5.10 or #5 and #6 in Figure 5.12) of that district were connected only by a single node. The results of this process are shown graphically in Figur es 5.10, 5.11 and 5.12.

Figure 5.10 Assignment Methods for CT #715

First Assignment Method for CT #715 - Bi-directional Grids

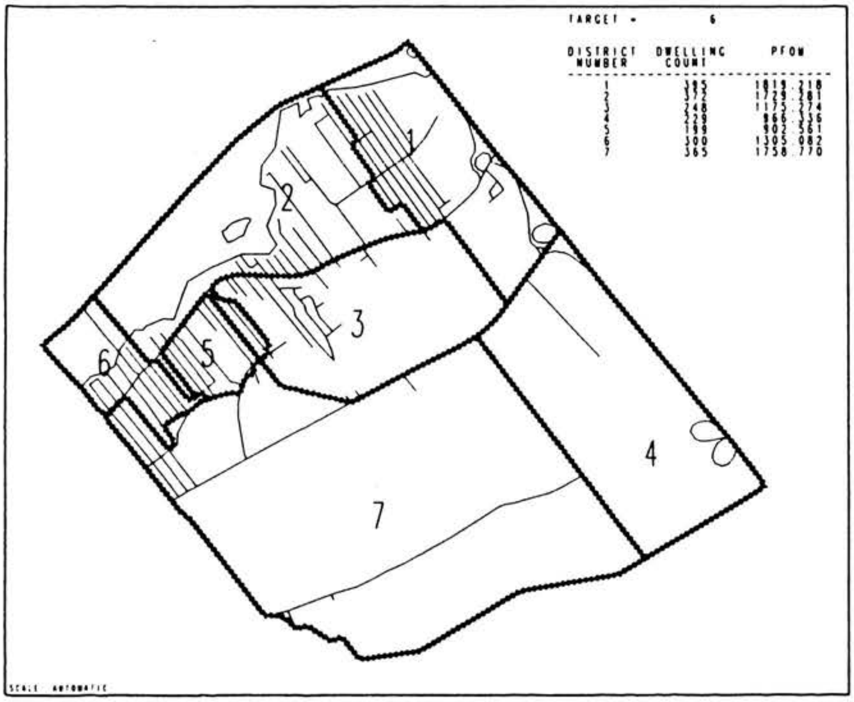


TARGET = 6

DISTRICT NUMBER	DWELLING COUNT
1	270
2	397
3	399
4	10
5	345
6	61
7	245
8	381

Figure 5.10 (Continued)

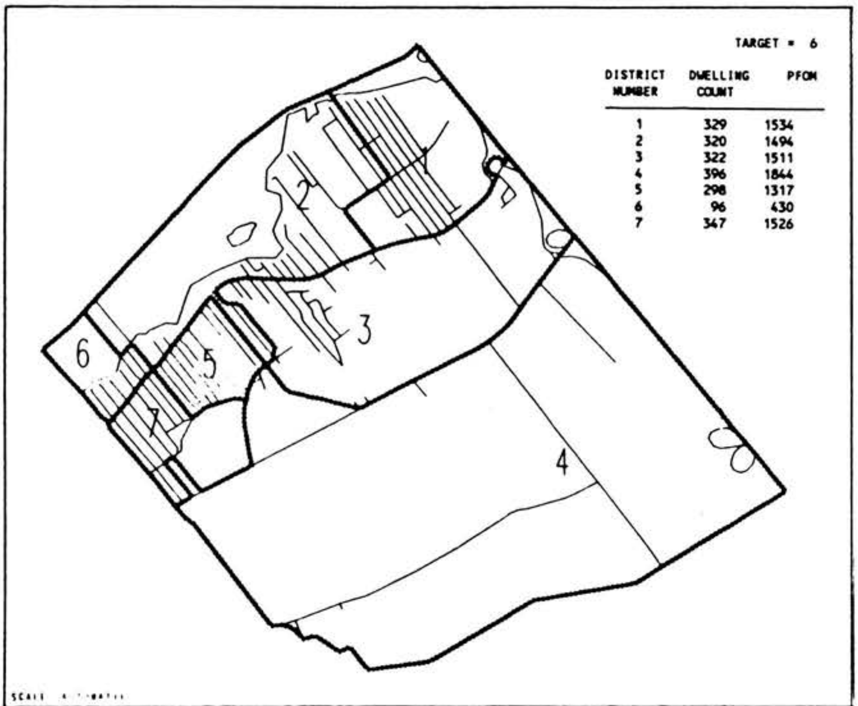
Second Assignment Method for CT #715 - Extrema-Based Seeds



TARGET = 6

DISTRICT NUMBER	DWELLING COUNT	PFOM
1	385	1819.218
2	372	1729.281
3	248	1175.274
4	229	966.336
5	199	902.561
6	300	1305.082
7	365	1758.770

Third Assignment Method for CT #715 - Sectors

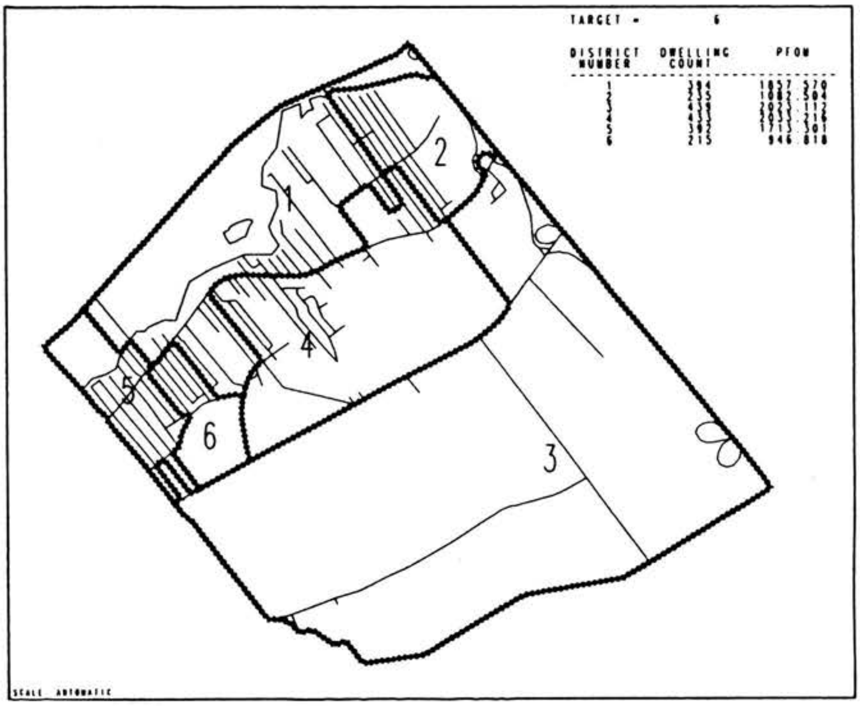


TARGET = 6

DISTRICT NUMBER	DWELLING COUNT	PFOM
1	329	1534
2	320	1494
3	322	1511
4	396	1844
5	298	1317
6	96	430
7	347	1526

Figure 5.11 Re-Assign Methods for CT #715

First Re-Assign Method for CT #715 - Sectors



Second Re-Assign Method for CT #715 - Extrema-Based Seeds

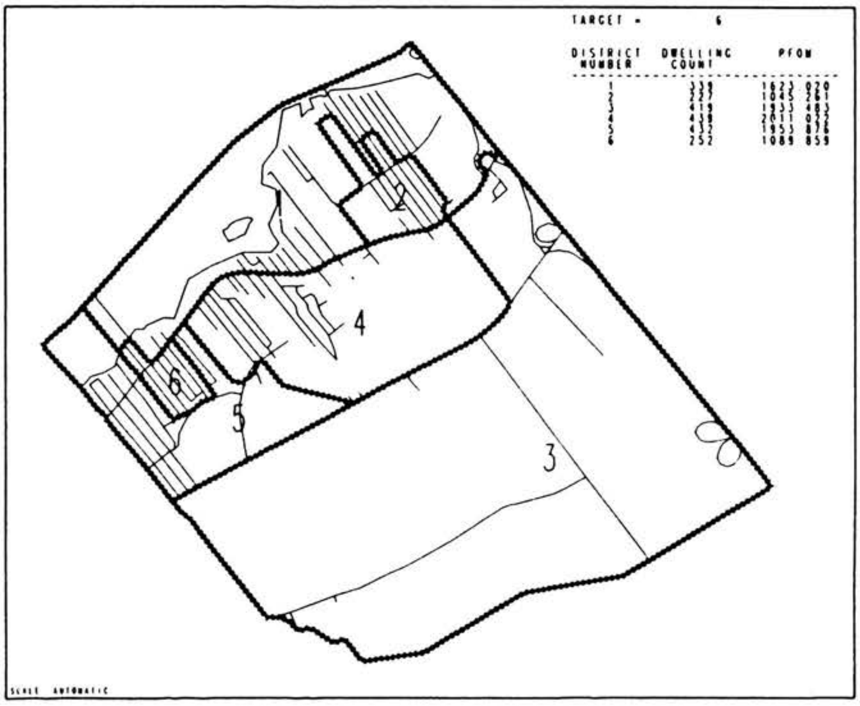
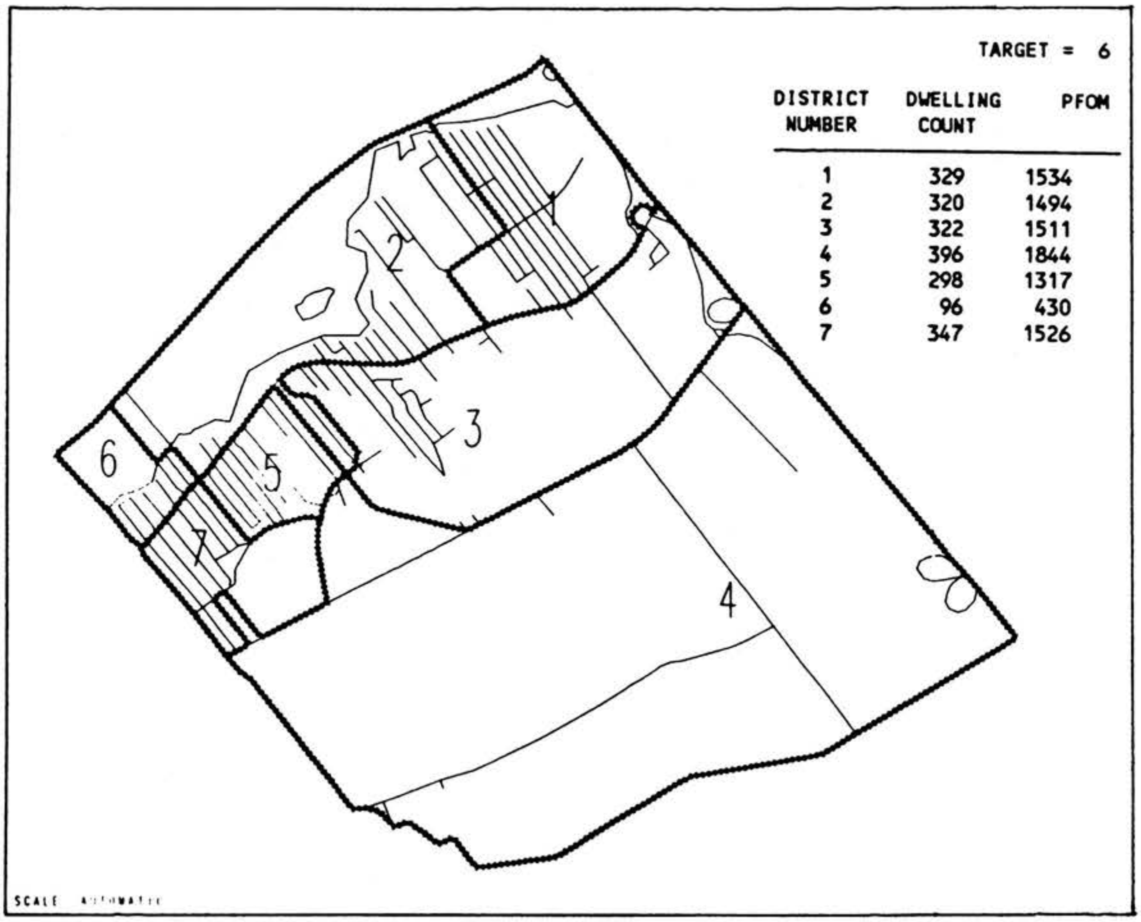


Figure 5.12 Adjust Method for CT #715

Adjust Method for CT #715 - Node Join Exemption on Sectors



5.6.4 Analysis Of Results On Case Study Areas

The results from this viability testing were very encouraging, if not conclusive because of the small sample, since the unit cost of producing a districting by an ASSIGN method was very low and the number of successful districtings (especially since a significant percentage of previous districtings could be reused) was quite high if discontiguities are not taken into consideration. Even more encouraging were the early results of testing using the new (and least expensive) ANNEX process which eliminated many of the discontiguities. The decision was taken, therefore, to incorporate the ANNEX process and to raise the standard for accepting a districting (i.e., discontiguous parts would be, henceforth, considered separate districts and assessed accordingly) and to incorporate the use of the ANNEX process on a larger (i.e., 'volume') test for the entire CSD of Laval.

5.7 THE LAVAL VIABILITY TEST (ON 63 CENSUS TRACT PARTS)

The full scale viability testing for Laval included considerations of cost and throughput since this testing served as the basis for a decision on whether or not to implement the results of this dissertation research as part of the 1991 Census. (See Statistics Canada, 1989, for an assessment of the results of that testing). The detailed summary of key results for each of the 63 CTPs in Laval (from over 600 test runs) is provided as a separate appendix (G). The main consideration in this section is on the quality of the districtings. The section includes a description of the measures that were developed to assess the quality of the districtings and a summary of the results of applying these measures across the manual and automated results for the CSD of Laval.

5.7.1 Comparative Quality Assessment

The quality of the districtings is represented by a series of quality measures of increasing complexity/sophistication. Each measure considers a different element of quality. Some measures compare the overall results to a given standard and are termed absolute measure s, and others compare the individual results to the actual, manually generated districtings utilized during the 1986 Census and are termed relative measures.

o Absolute Measures

1. Relationship to the Target Number of Districts.

Districtings generating fewer districts and especially those that can be combined into fewer workloads are considered, for operational purposes, to be of higher quality than those that do not. This is because they can result in savings in the staffing and the training processes and in the number of supervisors required. (However, actual enumeration costs are not altered dramatically because the Census Representatives are paid by the number of dwellings they enumerate.)

An index of quality for each districting alternative can be obtained by dividing the actual number of districts generated by the target number of districts for that CTP.

The index values in Table 5.9 were generated for the Laval Volume Test. The lower the value of this index, the better the result. Table 5.9 compares the actual results for the 1986 Census with dwelling based and PFOM based districtings for both the collection units and the actual workload units assigned to a Census Representative (Enumerator).

Table 5.9 Ratio of Actual or Generated Districts to Target

Dwelling-Based Approach Results	PFOM-Based Approach Results
Target # of districts = 263	Target # of districts = 276
Actual 1986 districts = 284	Actual 1986 districts = 284
Actual 1986 Census Districts Index = $\frac{284}{263}$ = 1.08	Actual 1986 Census Districts Index = $\frac{284}{276}$ = 1.03
Dwelling-based Autodistricting Index = $\frac{267}{263}$ = 1.02	PFOM-based Autodistricting Index = $\frac{278}{276}$ = 1.01
Actual 1986 Census Workload Index = $\frac{268}{263}$ = 1.02	Actual 1986 Census Workload Index = $\frac{268}{276}$ = 0.97
Autodistricting Workload Index = $\frac{244}{263}$ = 0.93	Autodistricting Workload Index = $\frac{255}{276}$ = 0.92

The dwelling based target number of districts for Laval is 263 and the PFOM based target is 276 districts. The actual districting of Laval for the 1986 Census used 284 districts (297 if "special" sub-blockface districts are included) which were later combined into 268 Census Representative workloads (or groups of small districts). The dwelling-based results from the autodistricting model generated 267 districts which were later combined manually into 244 workloads.

Therefore, in the case of the Volume Test for Laval, up to 24 fewer Census Representatives and, consequently, 2 fewer supervisors (commissioners) would have been required based on the use of the results of the dwelling based autodistricting approach.

The PFOM-based results generated 278 districts which were later combined manually into 255 workloads. Therefore, approximately 13 fewer Census Representatives and, consequently, 1 fewer supervisor would have been required based on the results of the PFOM based autodistricting approach.

Since the CSD of Laval represents approximately 1.6% of the total number of Census Tracts, the potential for additional savings in the data collection process at the national level are considerable in either case.

2. **Number of "Unsuccessful" Districtings.**

The evaluation function(s) used to control the autodistricting process can also be used to assess the end results for both manually and computer produced output. The elements of the evaluation function for both the dwelling based and the PFOM based approaches were detailed in Chapter 4.

An index for unsuccessful districtings of a given region (e.g., the CSD of Laval) is formed by dividing the total number of unsuccessful CTP districtings by the total number of CTPs to be districted (i.e., 63 CTPs).

The results of this type of analysis for each of the approaches applied during the Laval Volu me Test are shown in Table 5.10 and displayed in Figure 5.13 together with the results that would be obtained using the actual districts from the manual districtings for the 1981 and the 1986 Censuses.

Table 5.10 Unsuccessful Districtings Indices

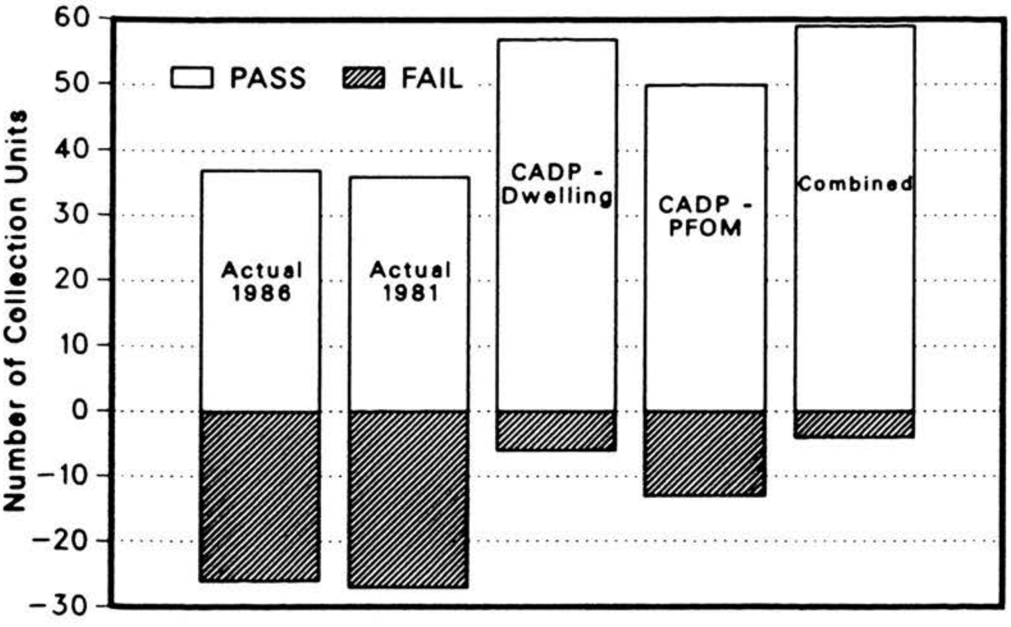
Evaluation Results From A Dwelling-Based Assessment			Evaluation Results From A PFOM-Based Assessment		
Case	Count	Index	Case	Count	Index
1981 Districts	27	0.43	1981 Districts	27	0.43
1986 Districts	26	0.41	1986 Districts	25	0.40
Autodistricting Based on Dwelling Counts	6	0.10	Autodistricting Based on Dwelling Counts	13	0.21
Autodistricting Based on PFOM Values	13	0.21	Autodistricting Based on PFOM Values	6	0.10
Combined Autodistricting	4	0.06	Combined Autodistricting	3	0.05

As might be expected, the performance of the two autodistricting approaches is relatively better when they are assessed by the method on which the approach is based. It is clear from Table 5.10 that, used alone (i.e., based on either dwellings or PFOM values) or in combination with re-using the previous districtings where appropriate, they perform exceedingly well. Based on the values of the resulting index, individually, they are 2 to 4 times more effective than the manual solutions based on this type of assessment of quality. That is, if viewed as a percentage failure rate, the manual approaches vary between 40% and 43% while the automated approaches vary between 10% and 21% if used individually or between 5% and 6% if used in combination with re-employing districtings from the previous (1981) Census.

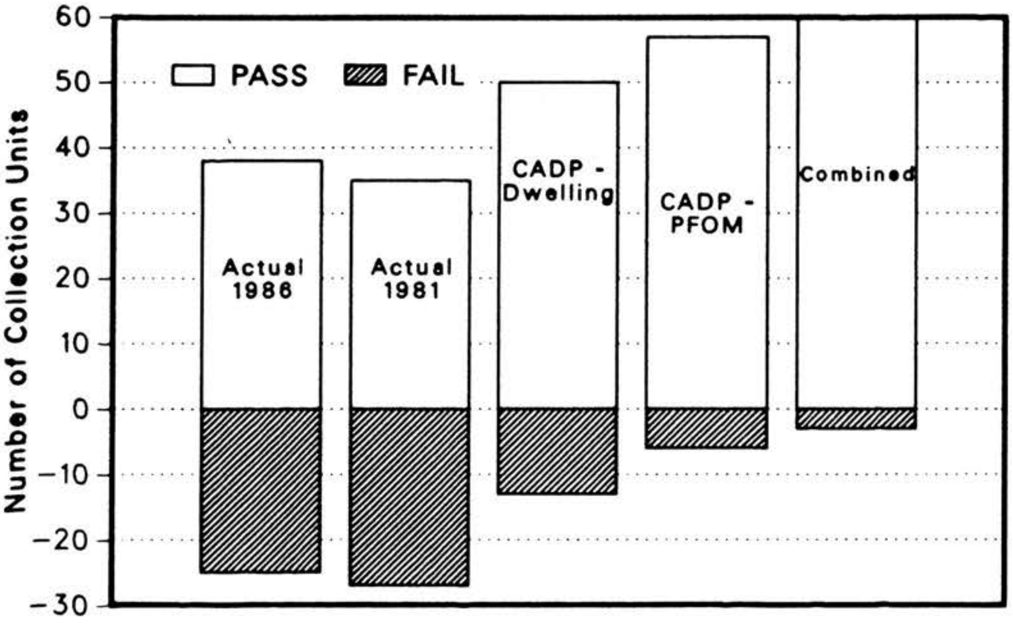
Figure 5.13 PASS/FAIL Analysis

CADP RESULTS

Dwelling Based Assessment



PFOM - Based Assessment



o Relative Quality Measures

The relative measures make a direct comparison between alternative districtings and the actual districting used for the 1986 Census on a case by case basis.

This relative quality is assessed by examining:

1. the ratio of PASS and FAIL results;
2. a combination of the number of districts generated and the average deviation value (relative to the target number of districts); and
3. the variation in the frequencies of the sizes of districts across the entire CSD of Laval.

The results for each of these types of analysis is presented in turn.

1. PASS/FAIL Analysis

A districting fails the evaluation function test if it exceeds the target number of districts or if the average deviation exceeds the specified range within an arbitrary number of autodistricting attempts. The number of attempts is based on pragmatic considerations that are primarily related to cost/benefit considerations (i.e., the cost of the automated approach would exceed the cost of the manual approach).

The method of assessing whether a districting passed or failed was described earlier (in section 5.2). A straightforward comparison, by CTP, between the actual, manually produced, 1986 Census results and those from districtings for the previous 1981 Census or from the automated system (run in 1988) can each result in four cases: pass/pass, pass/fail, fail/pass, and fail/fail. The results of this type of analysis applied to dwelling-based and PFOM-based aggregations, is shown, respectively, in Table 5.11a and 5.11b below. For example, in Table 5.11a, it is evident that, for dwelling based districtings, 37 of the 63 CTPs districts were passed for both the actual 1986 Census manual and the autodistricting procedures; 6 of the 63 were failed for both; and 20 of the 63 were passed for autodistricting and failed for 1986 Census manual districting. None of the districtings were passed for manual districting and failed for autodistricting. Thus, on a case by case basis, the manually produced actual 1986 Census districts passed in 37 of 57 or 65% of successful cases and the dwelling based autodistricting passed in 57 of 57 or 100% of such cases based on a dwelling based assessment.

Comparing individual, actual 1986 Census districts with PFOM based methods, there were 8 districtings which were unsuccessful for both the actual 1986 Census and autodistricted results. Thus, the PFOM-based autodistricting succeeded in 50 of 55 or 91% of successful cases while the actual districts succeeded in only 37 of 55 or 67% of such cases.

Table 5.11a Qualitative Performance Assessment Based On Dwelling Count Comparisons With Actual 1986 Districts

	1981 Districts MANUAL			Autodistricting Based On Dwelling Counts			Autodistricting Based On PFOM Values			Total Combined	
1986	Pass	Fail	1986	Pass	Fail	1986	Pass	Fail	1986	Pass	Fail
Pass	28	9	Pass	37	0	Pass	32	5	Pass	37	0
Fail	8	18	Fail	20	6	Fail	18	8	Fail	22	4

The values provided under the heading 'COMBINED' reflect the results obtained when the production system first checks the feasibility of re-employing the previous (i.e., 1981 Census) collection units and, only then, 'autodistricts' using either dwelling or PFOM-based approaches. This combined approach results in an **absolute** pass rate of 59/63 or 94% in the case of dwelling based districting, and a **relative** pass rate of 59 of 59 or 100%.

From Table 5.11b it is evident that the **absolute** pass rates for PFOM based districting were 57/63 or 90% for **autodistricting** and 38/63 or 60% for the manually produced, **actual** 1986 Census districts. The combined approach results in an **absolute** pass rate of 60/63 or 95%. The **relative** pass rates were 50/52 or 96% for the dwelling based approach (compared to 38/52 or 73% for the actual districts); 57/59 or 97% for the PFOM based approach (compared to 38/59 or 64% for the actual districts); and 60/60 or 100% for the combined approach (compared to 38/60 or 63% for the actual districts).

Table 5.11b Qualitative Performance Assessment Based On Figure Of Merit Comparisons With Actual 1986 Districts

	1981 Districts MANUAL			Autodistricting Based On Dwelling Counts			Autodistricting Based On PFOM Values			Total Combined	
1986	Pass	Fail	1986	Pass	Fail	1986	Pass	Fail	1986	Pass	Fail
Pass	31	7	Pass	36	2	Pass	36	2	Pass	38	0
Fail	5	20	Fail	14	11	Fail	21	4	Fail	22	3

A comparative index is generated by dividing the number of autodistricting successes by the number of successes for the actual 1986 Census districts as shown in Table 5.12.

Table 5.12 Success Rates For Previous Census And Autodistricting Districts (Compared with Actual Districts from the 1986 Census)

	Dwelling Approach	PFOM Approach
1981 Census Districts	$\frac{36}{37} = 0.97$	$\frac{36}{38} = 0.95$
Dwelling-Based Autodistricting	$\frac{57}{37} = 1.54$	$\frac{50}{38} = 1.32$
PFOM-Based Autodistricting	$\frac{50}{37} = 1.35$	$\frac{57}{38} = 1.50$
Combined Autodistricting	$\frac{59}{37} = 1.59$	$\frac{60}{38} = 1.58$

It is evident from Table 5.12 that autodistricting approaches (combined with the previous collection unit districts or individually) perform between 30 and 60 percent better than the actual manual districtings.

2. Combined Merit Index

In addition to assessing whether or not the alternative districting passes or fails and comparing the outcome to the result for the actual 1986 Census, comparisons can be made based on:

1. the resulting number of districts; and/or
2. the average deviation scores for dwelling-based (HH CADP), PFOM-based (PFOM CADP) and combined (SCORE1) approaches.
   (Note: both HH CADP and PFOM CADP results emulate pre-1986 Census conditions and are based on data from the previous 1981 Census.)

Districtings can be identical, (I), equivalent, (0), strictly better (from slightly better, +, through significantly better, + +, to considerably better, + + +) or worse (-,--,---), or better in some ways and worse in others (+/-) than the actual 1986 Census districts as summarized below in Table 5.13a for dwelling count based districting and in Table 5.13b for PFOM based districtings. **Weighted**, 'WTed', (i.e., +++=3, --=-2, etc.) and **unweighted**, 'UnWTed', totals (i.e., simple comparisons of the number of + CTPs versus - CTPs) are provided.

The column labelled 'I' in Tables 5.13a and 5.13b indicates the number of times that the generated district was identical to the actual 1986 districting. The column labelled '0' identifies the number of cases where both the average deviation and the number of districts generated were equivalent to the actual 1986 Census districtings. The column labelled '+/-' refers to results which are somewhat better (e.g., fewer districts) and somewhat worse (e.g., a larger average deviation) than the actual 1986 Census district. Columns with 1, 2, or 3 +'s or -'s refer to cases where there was a difference in the number of districts, the average deviation and/or the PASS/FAIL status of the districting. Bonus scores (to a maximum of +++ or ---) were given for significantly fewer districts. The results for all 63 CTPs are compiled in Appendix G and are too large and too intricate to report on here except in summary form. A score of 5P0.0 for the actual 1986 districting means that the average deviation was zero, the districting contained 5 districts and PASSed the evaluation function. However a comparative autodistricting result of 3P0.0 for the same CTP would be rated ++ since it has 2 fewer districts. Conversely, a score of 7F22.0 would mean that 7 districts were formed with an average deviation of 22.0. If compared to a value of 8F61.83, however, it would receive a rating of '++' because 7 is less than 8 and 22.0 is less than 61.83. (See Appendix G for additional details on the assessment methodology.)

SCORE1 refers to the combined result of first using 1981 districts as part of the autodistricting approach (if the previous districts prove acceptable). In all cases, the first acceptable result in sequence (previous districts, method #9, method #2, method #6), and not necessarily the best result, is used as the SCORE1 result (because the production process normally "stops" once a solution is found).

Table 5.13a Quantitative Performance Assessment Based on Dwelling Count Comparisons With Actual 1986 EAs

Previous and Autodistricting versus Actual Districts

METHOD	+++	++	+	0	I	+/-	-	--	---	Total	WTed Net	UnWTed Net
1981 EAs	9	1	4	1	26	8	4	5	5	63	4	0
HH CADP	21	1	12	7	6	0	16	0	0	63	77	18
PFOM CADP	18	3	7	5	4	0	19	2	5	63	31	2
SCORE1	21	2	3	2	27	1	6	1	0	63	62	19

Table 5.13b Quantitative Performance Assessment Based On Figure Of Merit Comparisons With Actual 1986 EAs

Previous and Autodistricting versus Actual Districts

METHOD	+++	++	+	0	I	+/-	-	--	---	Total	WTed Net	UnWTed Net
1981 EAs	5	1	6	13	6	14	9	4	5	63	- 9	- 6
HH CADP	11	3	14	8	4	8	13	1	1	63	35	13
PFOM CADP	17	2	17	7	4	2	11	2	1	63	53	19
SCORE1	17	4	10	12	17	5	8	0	1	63	58	14

In compiling the weighted and unweighted totals, the 'I', '0' and '+/-' categories were excluded.

The difference between the weighted and unweighted scores is generally considerable in both Table 5.13a and Table 5.13b. This is typically because when autodistricting is worse than the actual 1986 Census districting, it is only slightly worse. Conversely, when autodistricting is better, it is often very much better.

Looking first at the dwelling-based evaluation approach in Table 5.13a, all of the other districtings outperform the actual 1986 Census districts in achieving targets if a weighted total is considered. (This is because of generally better ratings for individual districts and was true in spite of the fact that there is one more 1981 Census districting that failed than 1986 districtings since a smaller number of districts were generated for that CTP.) If unweighted values are considered, the manual districtings from the previous (1981) Census are better than the 1986 Census districts as often as they are worse.

When the other factors included in the Partial Figure of Merit are taken into consideration (as recorded, in Table 5.13b), only the previous (1981) Census districts are rated inferior to the actual 1986 Census districts (as might be expected since the 1981 Census collection unit districts were based on 10 year old data).

The differences in the weighted scores for the PFOM-based assessment are not as pronounced as they are in the case of the dwelling-based (only) assessment. This is probably due to the fact that factors in addition to the estimated dwelling count distribution are considered by the districting specialists. Hence, their results will be better rated by the more complete considerations included in the PFOM based assessment. This view is also supported by the drop in the unweighted scores for dwelling (18 --> 13) and SCORE1 (19 --> 14) results between dwelling-based (Table 5.13a) and PFOM-based (Table 5.13b) assessment. Conversely, the relative rise in the weighted and unweighted scores for the PFOM approach in Table 5.13b (compared to Table 5.13a) is an indication that autodistricting is able to nearly maintain its comparative advantage (scores of 53 and 19 for the PFOM based approach versus 61 and 18 for the dwelling based approach) when a more complex and more appropriate (i.e., more complete) districting model is used.

3. Frequency Distribution Comparisons

A comparative analysis of the frequency distribution by 50 dwelling intervals (with the exception of the first two classes which are combined) of the dwelling counts for the individual districts was performed for each of the various districting approaches. This type of analysis of quality was also performed at the level of the Census Representative Workload.

The results for the Dwelling Based Districtings and Workload Assignments in comparison to the actual 1986 Census districts are summarized in Table 5.14 and Figure 5.14.

Similarly, this type of analysis was performed on intervals of 200 for the PFOM approach and these results are summarized in Table 5.15 and Figure 5.15.

Table 5.14 Dwelling-Based Autodistricting Results Compared With Actual 1986 Census Districts By Size of Unit

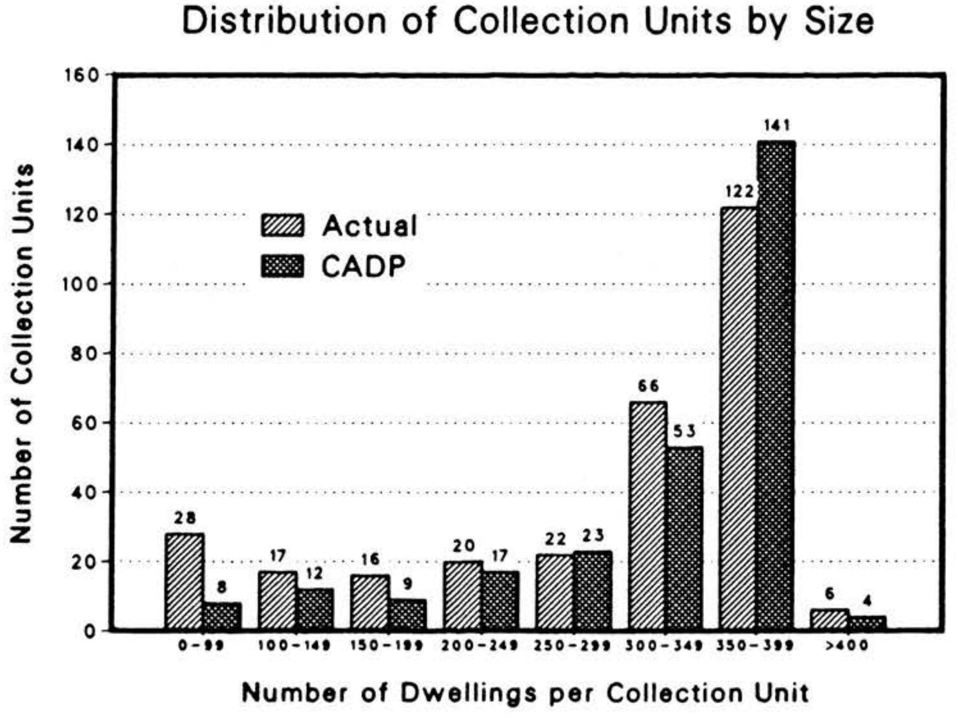
Dwellings	1986 Census Districts	CADP By Dwellings	1986 Census Workloads	CADP* Workloads
0- 49} 50- 99}	28	8	8	1
100-149	17	12	11	1
150-199	16	9	6	1
200-249	20	17	13	9
250-299	22	23	21	16
300-349	66	53	73	54
350-400**	122	141	128	153
>400	6	4	8	9
Totals	297	267	268	244

* Manually generated workloads from CADP districts.

** Target size range.

Figure 5.14 Autodistricting Versus Actual Results by Size (based on Dwelling Counts)

1986 COLLECTION UNITS CADP VS. ACTUAL



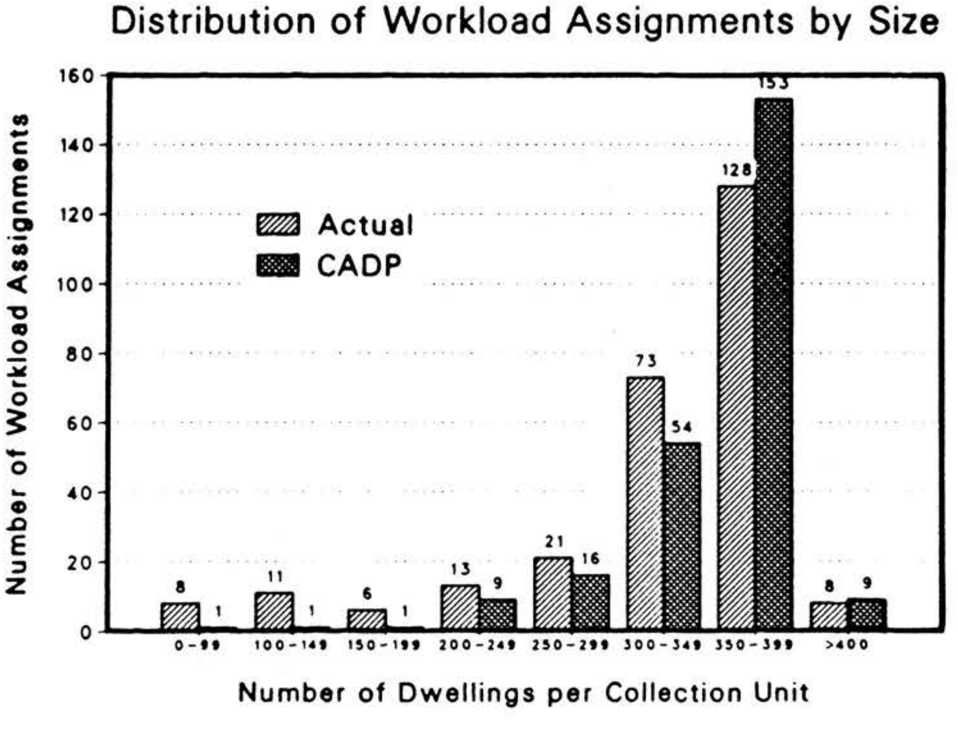


Table 5.15 PFOM-Based Autodistricting Results Compared With Actual 1986 Census Districts

PFOM Value	1986 Census Districts	CADP By PFOM	1986 Census Workloads	CADP* Workloads
0- 399	35	10	16	1
400- 599	12	7	6	1
600- 799	14	9	5	0
800- 999	17	14	11	3
1000-1199	16	18	19	13
1200-1399	43	49	39	50
1400-1599	114	132	118	143
1600-1799	42	37	48	44
>1800	4	0	6	0
Totals	297	276	268	255

* Manually produced workloads from autodistricting results.
The target range is 1350 - 1650.

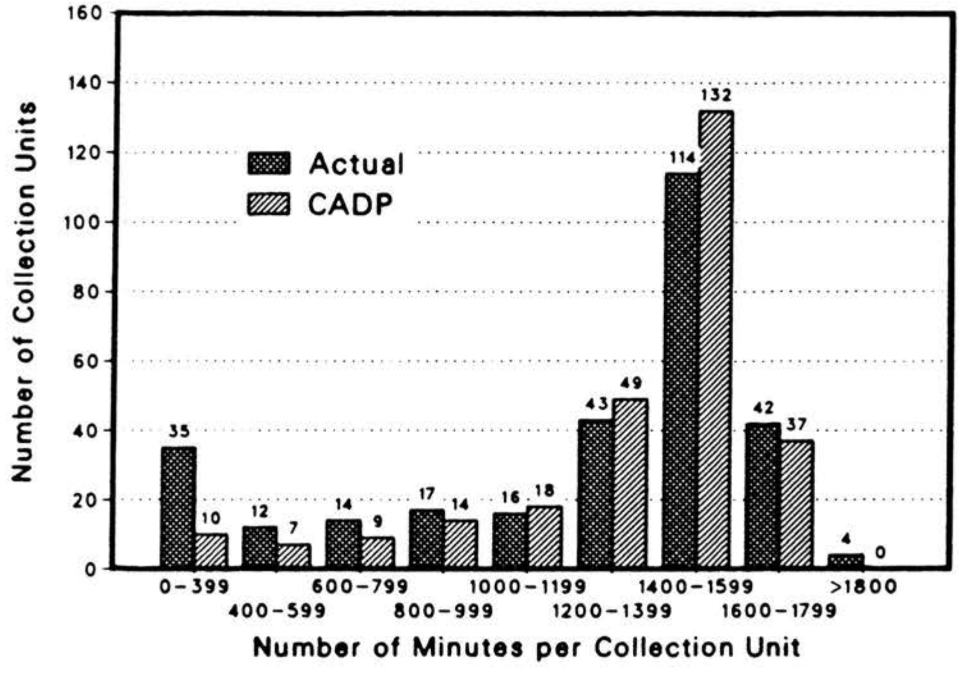
In general, the autodistricting results are more skewed towards the target size range. This is particularly true at the level of the workload assignments. The workload assignments are currently derived manually from both the manual districting and autodistricting results. (It is intended that the application of the autodistricting model be tested at the level of workloads during the 1991 Census since it is able to handle the separate sub-workloads using the topological transit component of the model).

It is important to recognize that these very positive results are the output of satisficing rather than an optimizing process. Indeed, they represent a comparison of the actual, manually-generated 1986 districts with the first acceptable results from the method selection process which is discussed further in the next section.

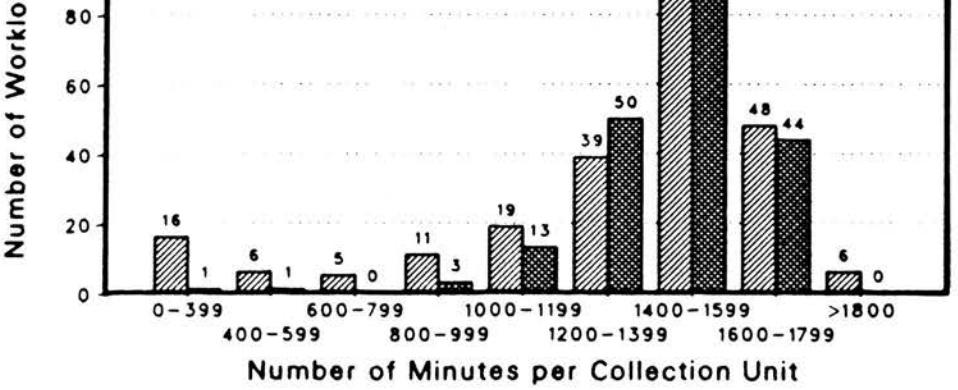
Figure 5.15 Autodistricting Versus Actual Results by Size (based on PFOM Values)

1986 COLLECTION UNITS CADP VS. ACTUAL

Distribution of Collection Units by Time Needed



Distribution of Workload Assignments by Time Needed



5.7.2 Results Of Testing The Method Selection Mechanisms

The method selection algorithms, designed by the author and tested with the assistance of the implementation team, were incorporated in the model primarily as an efficiency measure in the short term.

The goal was to use automated processes to reduce from 9 to 3 the number of attempted ASSIGN/ANNEX processes before moving on to the RE-ASSIGN and ADJUST processes.

Selecting 3 of 9 methods at random was expected to generate a 33% probability of selecting one of the three best methods for a given CTP. Tests that were conducted based on an initial decision table (see Statistics Canada, 1989) and on Table 4.5 (see Brockwell, 1988) -- which are both based, in part, on the number of districts to be generated -- failed to produce a result that was significantly better than could be expected from a random selection.

To assess the success rate of those two selection tables, it was necessary to produce and rank all nine methods for each of the Census Tract Parts (excluding those for which the previous Census districts would be re-employed). In producing these rankings, it was clear that a small subset of the methods (#9, #2 and #6) tended to be the most successful across the full range of cases. Consequently, it was decided to formalize their selection (in sequence -- #9, #2, and #6) for the initial version of the production system. Attempts to determine why these relatively straightforward/simple approaches (as compared to seed methods and bi-directional circular grids) were so successful were not conclusive (see Witiuk and Brockwell, 1990). Given the relatively small geographical areas involved, the relative homogeneity of settlement patterns and the generally small number of districts, perhaps the results are not too surprising. It is intended that the more complex methods be kept for future use and testing in cases that have higher variability.

However, in parallel, testing of improved method selection methods based on Tables 4.1, 4.2 and 4.3 continued.

Perhaps surprisingly, the most successful of the approaches is also the one that requires the least amount of information about the distributions within the individual Census Tracts. Indeed, the method selection based on the elongation ratio and the area ratio for the boundaries of the individual Census Tract boundaries (see Table 4.1) enjoyed the greatest success (Witiuk and Brockwell, 1990). Based on a sample of 24 CTPs, this approach was able to select 43 of 72, or 60%, of the best three methods. This was almost twice as likely as a random selection approach. Indeed, at least one of the top three methods was selected for all but one CTP. This is slightly better than the production system approach (i.e., selecting methods #9, #2 and

#6) which selected 40 of 72 or 56% of the best three methods for the same sample.

Post doctoral research will continue to evaluate alternative method selection algorithms with the long range objective of building a feedback mechanism that will automatically refine the ranges for the selection classes and the allocation of methods to classes.

5.7.3 Conclusions From The Viability Test

The volume test for the CSD of Laval has demonstrated conclusively, for each of the measures developed for objectively comparing the quality of manual and automated approaches, that autodistricting produces districts that are of generally higher quality than the manually generated actual 1986 Census districtings. Specifically, in terms of absolute measures of quality, the ratio of actual or generated districts to the target number of districts favoured autodistricting over the actual 1986 districts at the collection unit and at the workload level for both the dwelling based and the PFOM based assessment. That is, they were combined more effectively into suitable workloads than those that were produced by traditional methods. Further, the percentage of unsuccessful districtings was significantly lower for autodistricting (uniquely or combined with the previous districts) than for the actual 1986 districts.

Similarly, for the relative quality measures, the autodistricting results surpassed the manually generated 1986 Census in categorical (i.e., pass/fail) rates and for weighted and unweighted combined merit index values (which includes the number of districts and the average coefficient of deviation values). The reduction in the number of Census Representatives resulting from autodistricting is an important operational measure of quality as it can lead to significant savings in recruitment, training and supervision costs for Census data collection.

Finally, the autodistricting results, when examined from the point of view of the frequency distributions at the collection unit and at workload levels outdistance the manually generated actual 1986 Census districts.

Outside of the scope of this research activity, the actual cost of districting operations was determined (Statistics Canada, 1989) to be significantly lower (on the order of $250,000 less if 150 FEDs were to be autodistricted for the 1991 Census).

5.8 LIMITATIONS

This section documents the research limitations in:

1. the selection of case study areas;
2. the nature of the tests;
3. comparing manual versus automated results;
4. the AMF specific enhancement techniques (i.e., block bonding methods); and
5. the utility of the produced capacity.

5.8.1 Limitations In The Selection Of Case Studies

A more objective method of selecting CT test sites is to randomly select candidate sites from sets of small, medium, and large CTPs that are stratified by dwelling densities and street pattern types.

Given the pragmatic constraints of data availability and the desirability of having linguistic diversity, and historical geographic stability, it is believed that no significant improvement in the representativeness of selected sites would have been realized with a less arbitrary approach.

In relation to data availability, it should be pointed out that the production sequence for delineating collection units and revising basic cartographic information was also not random. Indeed, for the 1986 Census, considerable care was taken in sequencing the earliest centres to provide a suitably varied (i.e., challenging) framework for testing and implementing the computer-assisted collection mapping (CACM) system under development at that time.

Here also, a careful balance between selecting representative centres and selecting sites for which timely supporting cartographic information would be available from the local authorities was necessary.

As it turned out, one of the selected CTPs was one of the most difficult cases in all of Laval and was one of 6 (of 63) that required adjustment. Thus, the percentage of unsuccessful districtings was higher (25%) for the sample of 4 CTPs than was the case for the entire CSD (10.5%).

5.8.2 Limitations In The Nature Of The Tests

This section discusses the limitations in the nature of testing of approaches not based on dwelling counts or the partial figure of merit, of the topological transit component of the models and of the method selection procedures.

1. **Testing of approaches not based on dwelling counts**

For practical reasons the focus of the current testing was on the traditional dwelling count aggregation approach and the new 'elapsed time-based' figure of merit approach. The testing of the other approaches: block count, perimetre, surface area, and regular geometric partitionings was adequate to ensure that the model functioned as expected. These later approaches are intended to be employed only in those areas where dwelling counts are not readily available (such as for digital cartographic files for areas outside of the coverage of the Area Master File Program available from external sources such as National and Provincial Mapping Agencies). Hence, in-depth testing of these other approaches was not seen as urgent, nor critical to the success of this dissertation.

2. **Testing of the 'topological transit' component of the model**

The 'topological transit' component of the model is a useful instrument for dealing with natural and "person-made" barriers to accessibility. Unfortunately, the current cartographic data base does not carry sufficient information to permit a full implementation of this concept. The structure of these files is currently under review and it is expected that their evolution will address these current shortcomings. For the moment, the use of topological proximity is employed to an adequate degree in the ANNEX and RE-ASSIGN processes of the model.

Some controversy remains amongst districting specialists and field operations staff as to the connectivity of blocks sharing a common intersection point but no sides. Future field testing should address the Census Representatives' ability to handle collection units of that type.

3. **Testing of The Method Selection Procedure**

The initial testing of the method selection algorithm was based primarily on the results of the pilot study areas. Brockwell (1988) concluded that "the results indicate that there was a small difference in favour of the selection algorithm" over a random selection of methods. However, the pilot study did not provide a large enough sample to generate a confidence interval of greater than 90%. Thus, the method selection algorithm was re-tested for the 24 Census Tracts Parts in Laval that required redistricting. Based on the larger sample, it was later concluded that, while the method selection procedure was much more successful than a random selection, the correlation was tied more to the performance of the best methods than to the value of the coefficient of variation. Therefore, it was decided to retroactively apply additional selection tests based on Tables 4.1, 4.2 and 4.3 to the 24 CTPs. As mentioned, the most effective of these has proven to be Table 4.1 which was almost twice as successful as random selections.

The production process currently in place by-passes the method selection algorithm and proceeds sequentially through the three methods that performed best during the Laval Volume Test. The limitations of this pragmatic approach stem from the practical necessity to establish an interim approach for the production system. As the model is employed for the 1996 Census, further testing and analysis should be undertaken to determine the advantages of re-implementing the automated selection algorithm. Of particular interest should be an analysis of the impact on the selection of methods operating on the CTP boundaries with a fixed shape (and characteristic distributions) but with different geometrical orientations to see which measures are the impacted the least by the orientation of the street network pattern.

5.8.3 Limitations In Comparing Manual and Automated Results

The comparison between manual and automated results was based on desk exercises at Statistics Canada headquarters in Ottawa. Initially, the fact that the pilot study assumed that all block splits would be handled by districting specialists using the ADJUST process and the fact that all but one of the test sites required at least one block to be split, precluded the possibility of a head-to-head, blind comparison in the field of the two results during the pilot study stage since the lack of block splits permits all automated results to be easily discriminated from manual results and could significantly lower the objectivity of the results.

Later, the incorporation of block splits from previous census districtings in the Volume Test study for all of Laval would have permitted such testing to be conducted. However, the general acceptance of the "operational adequacy" of the automated solutions by all participants by the time such testing was possible, precluded the requirement to determine if, on a case by case basis, they would also be perceived as being of superior quality by Regional Office staff responsible for conducting the field checks.

The general difficulty of comparing manual and automated results was evident from informal assessments of alternative districtings (manual and automated) conducted at headquarters, early in the second stage. Even here, there was not a consensus between managers, coordinators, supervisors or districting specialists as to which of several acceptable districtings was the best. This lack of authoritative and generally held perspective on quality is made even more complicated when elevated to the level of comparisons across an entire CSD.

Global comparisons between the ratio of acceptable to unacceptable districtings would have to be classified in great detail according to the reasons for unacceptability. Specific districtings may have been rejected because of:

1. recent growth in the number of dwellings (unknown to the manual and automated districting processes);

2. recent introduction of new accessibility impediments/improvements (e.g., new streets, freeways, closed bridges, etc.);

3. local knowledge of the parts of the Census Tract that are designated as "difficult to enumerate";

4. recent shifts in the linguistic character of a neighbourhood; or

5. the decision to try to accommodate local priorities for historical contiguity (in spite of possible inefficiency) or for conformance with local zoning patterns (e.g., wards, parishes, etc.).

Similarly, preferences between acceptable districting on a case by case basis were likely to be based on subjective criteria which, relative to the objective specifications, were not defensible such as:

1. the relative elongation of alternative districtings, though the distances to be travelled are comparable;

2. the desirability of splitting neighbourhood along better known (locally) or more travelled streets (e.g., with fewer stop signs);

3. variable individual preferences to use certain feature types (e.g., railroads) over other (rivers, hydro-electric transmission lines, etc.) as non-street collection unit boundaries;

4. the local preferences to include blocks with certain types of land use (e.g., parks, schools, cemeteries, etc.) with specific neighbouring areas; and

5. local knowledge on the preferred manner in which to split a densely settled block (e.g., location and size of specific dwellings).

Given the size of the sample that would be necessary and the degree of segmentation and/or control of the objective and subjective factors influencing individual judgements, it was felt that little would be gained from a head to head, blind test in the field of manual in light of the fact that automated districting viewed as operationally adequate by headquarters staff.

5.8.4 Limitations In Block Bonding Methods

Limitations in time and in the content of the current Area Master Files precluded the implementation and testing of more sophisticated methods for the pre-grouping of blocks with no dwellings with their neighbours, the early grouping of fully embedded blocks with their "parent block" and the post-grouping of blocks with "natural affinities" (such as islands) or with flexibility in association with neighbouring blocks (e.g., two large subdistricts neighbouring two small subdistricts can be better combined if increased searching and checking is introduced). In future versions of the Area Master Files, the blocks with no dwellings are likely to be more accurately classified (boulevards, traffic islands, recreational and/or industrial parks, etc.). This will allow for implementation and testing of improved 'block bonding' methods. Similarly, the adding of explicit topology to the AMFs will facilitate grouping neighbours whose workloads "exactly" complement each other. (This would be like applying the ANNEX process to large blocks before using an ASSIGN process.)

5.8.5 Limitations In The Utility Of The Autodistricting Capacity

The autodistricting capacity that has been implemented based on the concepts, methods and specifications developed as part of this dissertation has proven to be successful in generating operationally adequate districts without human intervention in 57 of 63 test cases. Of the remaining 6 test cases, all but one, have been judged by Field Operations staff as acceptable as traditional districtings for the same areas. The remaining test site was easily modified to meet the standard using the "ADJUST" module of the program.

5.9 CONCLUDING COMMENTS ON TESTING AND RESULTS

This chapter has described the kinds of testing necessary to validate the family of districting approaches and models that has been implemented as a result of this research effort. Although the testing program was intricate and time-consuming, it has served not only to validate the theoretical components of the model, but also demonstrated the practical value and flexibility of the approach.

While there were pragmatic limitations on the amount of testing that could be undertaken prior to submitting this dissertation, the testing and enhancement of the model and the implemented system will continue. Because of the increased quality, throughput and efficiency it affords, the model has been adopted for the 1991 Census.

Systematic consideration has been given to each of the 20 characteristics identified in the statement of the problem (Chapter 2) as is shown in Table 4.10. When a specific characteristic (or objective criterion) is deterministically assured by invoking established human or GIS processes, the method of implementation is termed 'procedural'. When a specific characteristic is 'striven for' through iteratively employing the heuristics developed as part of this research, the method of implementation is termed 'methodological'.

Table 5.16 CHARACTERISTIC IMPLEMENTATION TABLE

CHARACTERISTICS	METHOD OF IMPLEMENTATION
Completeness	Procedural (verified by overlay)
Uniqueness	Procedural (verified by overlay)
Hierarchy conformant	Procedural by polygon overlay
Single methodology	Procedural by selection of CTs
Agricultural limits respected	Procedural by selection of CTs
Historical Continuity	Procedural (selection method 0)
Minimum number of zones	Procedural (ND calculation)
Respect visible features	Procedural (block formation)
Facilitate Accessibility	Methodological (Transit Table)
Consider Supervisor Workloads	Procedural (Grouping Residuals)
Respect Workload Limits	Methodological (PFOM component)
Respect Collectives	Procedural (Sub-blockface FILTER)
Respect Blockfaces	Procedural (Blockface FILTERING)
Respect Blocks	Procedural (Block FILTERING)
Respect Contiguity	Methodological (ANNEX,RE-ASSIGNS)
Respect Linguistic Groupings	Procedural (ADJUST Process)
Minimize route length	Methodological (PFOM component)
Minimize route start distance	Methodological (FOM component)
Maximize homogeneity	Procedural (Definition of a CT)
Strive for compact shapes	Methodological (PFOM Component)

Thus, fourteen of the twenty characteristics of an "optimal" districting are achieved (in the parts of the country covered by both the Census Tract Program and Area Master Files) by fixed procedures incorporating either customized FILTERING processes or standardized Geographical Information System techniques. Six of the twenty characteristics are 'striven for' using the ASSIGN, ANNEX, and RE-ASSIGN processes that have been developed as part of this research and implemented as a flexible tool kit of methods.

Of these twenty characteristics, two require additional discussion at this time, namely, respecting **linguistic groupings** and striving for **compact shapes**. It is also important to clarify the sense in which characteristics that have been classified as methodological are "optimized".

Linguistic Considerations

The current version of the model deals with linguistic considerations (i.e., the minimization of the number of bilingual districts) by using standard ARC/Info mapping capabilities to identify collection units that offer reasonable prospects for realignment in light of the desire to respect linguistic groupings. Specifically, all collection units that are identified as bilingual and have a minority official language component of less than 20% are flagged (as a shading pattern on a standard choropleth map) for possible redistricting by the districting staff using the ADJUST process after the autodistricting process is complete. Reducing the number of bilingual districts in this way is straightforward and facilitates the staffing process since fewer bilingual Census Representatives are required.

Compact Shapes

It was argued in Chapter 3 that **compact shapes**, though a stated criterion for optimal collection unit districts, was of minor importance relative to characteristics (such as workload, route length, etc.) that are more directly related to the actual enumeration effort as shown in Figure 4.4. Without over-emphasizing this characteristic (e.g., by making one of the standard measures of compactness an explicit element of the multi-component objective function), the compactness of collection unit shapes is 'striven for' in the following ways:

1. the ASSIGN processes combine (by both decomposition and composition) **block** centroids that are geographically near to one another;

2. the ANNEX process combines subdistricts that are topological neighbours;

3. the RE-ASSIGN process combines **blocks** that are topological neighbours and gives precedence to the geographically nearest **blocks** (i.e., based on their centroids) if the

characteristic values are equivalent; and

4. the ADJUST process permits the districting specialist to improve the "optimization" of the districtings by further considering compactness. However, experience to date has found that most districtings are operationally adequate from the viewpoint of compactness without involving to such ADJUST processes (i.e., visual inspection indicates that they are typically fairly compact).

"Optimization" of Methodological Characteristics

Since, as discussed in chapter 2, an exhaustive optimization approach has been shown to be computationally impractical for more than 40 units, the practical alternative was to heuristically generate alternative solutions until a 'satisfactory' solution was found. This latter approach is the one that was pursued in this research.

A solution was typically considered satisfactory if the sum of the deviation of the dwelling counts, PFOMs, etc., for individual districts is less than 10% of the target value times the number of districts. (Or, said another way, the actual average deviation is within 10% of the target value.)

The heuristic optimization process attempts to "minimize" (or reduce to a manageable number) the total number of districts, while minimizing the weighted deviation of the individual districts from the target value. The deviations above the upper bound for the dwelling count and PFOM based approaches are "geometrically weighted" to take into account the more serious nature of oversized districts in comparison to undersized districts.

The final chapter reviews the research contributions, summarizes the principal findings and provides some closing remarks.

CHAPTER 6

CONCLUSIONS

6.1 RESEARCH CONTRIBUTIONS

Computer-assisted districting is an important tool for an essential process in geocartographics and draws upon fundamental elements of regional geography, general cartography, set theoretic and graphic theoretic mathematics, and combinatorial computer science. The application of computer-assisted districting capacities to the creation of collection unit districts for the 1986 Census of Canada offered a number of practical as well as academic challenges. From the practical point of view, this research addressed the question: "Is it feasible and cost/effective to create collection unit districts by semi-automated means?"

An evaluation of the prototype autodistricting capacity, involving both the end users of the capacity and the products produced, has found that collection unit districts can be generated (with very little human intervention) for the next Census at an estimated savings of 8 person years or $200,000 (if collection units for 150 Federal Electoral Districts are produced) [Statistics Canada, 1989]. Based on these findings, the decision was taken to use the production autodistricting capacity to district approximately 75 Federal Electoral Districts (yielding about 10,000 collection units out of a total of about 40,000) for the 1991 Census. (This production undertaking was successfully completed in February 1990.)

From the academic perspective, the research investigated the quality, utility and efficiency of alternative districting schemes. More theoretical questions involving the computability, the ease of integration, and the general applicability of the alternative approaches were also discussed.

The objective determination of the **absolute** and **relative** quality of collection unit districtings is an important contribution of this research. Prior to this research, districtings were subjectively assessed, at Statistics Canada and at other census agencies, as simply adequate or not.

This research developed a model and methodology to effectively produce (with minimal human intervention) Census Collection Unit Districtings of comparable or superior quality to those generated by traditional manual approaches. Their development led to improved insights into the tradeoffs that can and must be made between the size and number of districts and, at the level of the workload, the relative importance of having contiguous subdistricts.

A computer-based implementation of the model has been designed and, with the assistance of staff at Statistics Canada, implemented and tested. It cost-effectively generates districts for the collection of the Canadian Census that meet or exceed the quality standards of traditional methods. This implementation of the model provided a fertile environment for experimentation and for empirically establishing appropriate tolerances and ranges for algorithmic decision making. It also provided feedback on alternative approaches for resolving special cases (such as the collinearity of "block" centroids) and for deciding on the appropriate degree of look-ahead in ASSIGN methods.

In addition to the establishment of an operational autodistricting capacity that is quick, inexpensive, easy to use and that was (enhanced and then) employed for the 1991 Census of Population and Housing, this research has also provided a basis for further application to other field collection activities, and for further research into more sophisticated approaches (involving, for instance, greater use of expert systems) that will operate cost effectively in the computing environments of the future.

Because of the design approach taken, the tool kit provides a specialized set of techniques that can be used in combination and applied to a variety of similar problems. The intent was achieved of having a variety of approaches (ASSIGN, ANNEX and RE-ASSIGN) that can be used alone or in combination to keep to an acceptable level the need to employ the interactive ADJUST process and to maximize the level of flexibility.

This multi-stage, multi-component model can be adapted to apply, to a greater or lesser extent, to the broad range of districting applications identified in Chapter 1, namely:

1. survey taking;
2. mail or flyer delivery;
3. garbage collection areas [Berlin, 1974];
4. meter reading zones;
5. police surveillance areas; and
6. electoral register districts.

Each of these applications has an areal coverage component that has a significance comparable to optimization criteria such as route minimization and therefore a strictly optimum solution is not mandatory.

Indeed, the autodistricting prototype system developed during this research is being evaluated currently for possible application to the re-design of the sample frame for the Labour Force

Survey in Canada and for creating operational districts for maintaining provincial hydro-electric infrastructure.

One of the purposes of the research was to develop and document a better understanding of the (formerly undocumented) traditional districting methods. This research contributes a formalization of a former craft into a more scientific and measurable process.

Another purpose of the model was to improve the quality of work for districting specialists by automating labour intensive operations while producing operationally adequate districts in a much faster and cost-effective manner.

The successful approach taken in this research combined and extended elements of many sophisticated geographical information handling systems and districting methods in a novel and hybrid manner. That is, 'global' operators such as the ASSIGN processes that decompose (disaggregation) or compose (aggregation) the space into districts are combined with 'local' (i.e., REASSIGN) and 'intermediate' (i.e., ANNEX) operators which form districts based on topological adjacency and proximity to "core" seed locations.

Two completely new methods (ANNEX and RE-ASSIGN) were designed and employed to great advantage to improve the overall level of success for the CSD of Laval from about 59% to over 90% (or 98% if traditional "exemption" practices are taken into consideration).

6.2 SUMMARY OF PRINCIPAL FINDINGS

The rudimentary tools for interactively forming or reforming aggregations of blocks have existed for over a decade. However, until the completion of this research effort, there was a real doubt as to whether or not an almost fully automated capacity for creating Census collection units -- able to generate comparable or superior solutions to traditional methods -- was possible without resorting to extensive and costly human intervention.

Economic and coverage considerations ruled out the use of computationally intense alternatives that tend to reduce the problem to a highly simplified state in order to compute an "optimum" solution. Financial constraints ruled out the use of recent expert systems technology (tool kits and shells) which may not as yet be able to support such a complex application.

Pragmatics tempered the extensive use of expert systems techniques in the design of the system. Pragmatics further tempered their use as part of the prototype implementation. For example, the method selection component of the model was usurped in favour of a sequential selection of the most successful techniques based on empirical evidence.

The simple processes within the ASSIGN module outperformed more complicated techniques with unidirectional rectangular and circular grids (or sectors) outdistancing all other methods. Bidirectional grids and the novel extrema-based seeds methods also performed well.

The RE-ASSIGN method was constructed to ensure contiguity and to account for travel impediments while grouping individual "blocks". The primary data source, the Area Master File, proved inadequate to the demands of the "topological distance" component of the model, and hence, this element of the figure of merit, though implemented, was not fully exploited.

Further, the extension of pre- and post-districting 'bonding' of empty blocks into the ANNEX function for combining adjacent groups of non-empty blocks proved very effective and greatly reduced the need to use the more costly RE-ASSIGN function for the CSD of Laval.

The model has been developed for eventual application throughout Canada (i.e., with or without attribute information on the distribution of dwellings, linguistic groups, etc.). However, the focus for this implementation and testing has been on the use of the model in urban areas with over 50,000 population, where related block-face geocoded Census information and Area Master Files are available.

The testing compared results of the model with results produced by traditional means for the 1986 Canadian Census of Population and Housing.

In comparison, there was a 98% success rate without intervention since only one Census Tract Part out of 63 (a 1.6% sample of the national coverage) required the use of the interactive ADJUST process and since 5 of the 6 CTPs that failed the automatic assessment were granted traditional exemptions from the standard.

6.3 DISCUSSION OF LIMITATIONS IDENTIFIED IN CHAPTER 5

The amount and kinds of testing undertaken to assess the performance of this districting model and tool kit was limited by practical constraints. However, results from the production work for the 1991 Census, (which are currently being documented by production staff), demonstrate that the sample, though small and non-random was adequate for effective management decision making and relatively representative in terms of providing reasonable estimates of expected quality, cost and throughput.

The testing of the approaches not based on dwelling counts has been adequate to establish a relatively high level of confidence that the model/tool kit can be used for applications which do not have detailed dwelling distribution counts. At the other extreme, the high level of success of the dwelling based districting approach and the simultaneous exposure to the Figure of Merit based assessment statistics has created an environment whereby it is highly likely that the PFOM approach will be the approach of preference for the 1996 Census.

The limitations in the structure of the existing digital cartographic data files which inhibited the full exploitation of the topological transit component of the model are being investigated for a variety of practical reasons including the clearer understanding of the value of a complete, consistent and fully integrated data base, not only for districting but for several post-districting operations as well.

The degree of testing of the method selection procedure far exceed the practical requirement of the production process once it was determined that 3 of the 9 methods were able to resolve a very high proportion of the districting cases. From an academic perspective, there remains unresolved a considerable challenge. The results of the current level of analysis are being documented at present for future presentation and publication [Witiuk and Brockwell, 1990]. One of the useful concepts from the literature being assessed as a possible alternative selection index is the notion of spatial entropy. [Medvedkov, 1966; Batty, 1972]

The comparison of the quality of manual and automated districting results is fraught with difficulties unless undertaken in a completely controlled environment (i.e., same sites, criteria, timeframes, and source materials). The 1991 Census production workload provides a suitable framework for making such comparisons (perhaps by redoing a centre that was manually districted to assess relative cost, quality and throughput). Such a test could be expanded to an assessment in the field of the relative degree of operational adequacy of the results of the two production processes. Given the seemingly broad range of results considered operationally adequate, one wonders how practically useful such a field test would be since timeliness and cost per unit likely will be the determining factors in choosing between two districtings that are both operationally adequate (even if one is clearly superior in terms of quality). From an academic perspective it may be worthwhile to determine with greater precision what factors lead to the perception that one districting is better than another (even if they are equivalent from the viewpoint of an objective function). Such an assessment may also uncover additional data elements required of the data base or data base structure. Such elements may be particularly useful in 'bonding' together selected blocks (e.g., parks, traffic islands, boulevard dividers, or islands) with particular neighbouring blocks to facilitate geometric compositions or decompositions.

6.4 CLOSING REMARKS

Computerization of fundamental operations of a census began in 1890 with invention of the punch card by Dr. Herman Hollerith [Reid-Green, 1989].

A major step forward was made with the construction and implementation of the first commercial computer, the UNIVAC I for performing the basic tabulations of the 1950 Census of the USA. Each succeeding decade has witnessed the gradual introduction of more and more complex computer processes:

- the 1960's - automated geocoding and small area retrieval [Fellegi and Weldon, 1967];
- the 1970's - statistical analysis, classification and mapping [Broome and Witiuk, 1980]; and
- the 1980's - reference and collection mapping [Bradley, 1981].

In Canada, automating Census operations has proven typically to be more difficult and time consuming than originally estimated. This was particularly true for automating the districting function due to:

1. the initial lack of documentation on procedures and the lack of quantitative standards;
2. the intricacy of the relationships between competing criteria and objectives;
3. the incompatibility of data sets from successive censuses and the large volumes of data required;
4. the need for sophisticated geographic data handling (especially polygon overlay) and presentation (e.g., generating maps suitable for data collection) capabilities;
5. the anomalies presented by "real world" case studies (e.g., sub-districts sharing a common node); and
6. the massive amount of testing required for tool kit types of solutions.

It has taken longer than expected and has required substantial amounts of human and computational resources to ensure that the testing was adequate and that the choices amongst numerous alternatives were reasonable. Never-the-less, it has been worthwhile because of the increased understanding, documentation and co-operation that has already been achieved.

It can now be confirmed that, as a direct product of the research results presented in this dissertation, automated collection unit districting has been added to the list of automated Census processes in Canada in the 1990's, and can be expected to be used in other countries as well before the decade is complete.

Additionally, evaluation studies are under way to assess the viability of using this decision support system to generate: (a) "clusters" for collecting the results of the Canadian Labour Force Survey; (b) "weighting areas" for attributing the characteristics from the 1 in 5 sample (who receive the long form of the Census questionnaire) to the remainder of the Canadian population (who receive the short form of the questionnaire); and (c) "tiles" for storing groups of National Topographic Series (NTS) digital map sheets in a cartographic data base for the entire country.

Given the general utility of the autodistricting process and the power and flexibility of the implemented model, it can be expected that the results of this research will receive widespread use across a broad range of applications. Indeed, while the first of the three applications listed above was anticipated (along with applications such as creating polling zones, meter reading areas, or canvasser districts), the use of the model to meet internal operational requirements such as creating "weighting areas" or "tiles" was not expected. Undoubtedly, other such applications will surface as news spreads of the availability and effectiveness of the new capacity.

REFERENCES

Abler, R., Adams, J.S., and Gould, P.R., 1972: **Spatial Organization: the Geographer's View of the World.** Prentice/Hall International, Inc., London, UK.

Batty, M., 1972: "Entropy and spatial geometry", **Area,** Volume 4, p. 230 - 236.

Berlin, G.N., 1974: "Computerized Residential Refuse Districting", **Journal of the Planning and Development Division,** American Society of Civil Engineers, Volume 100, Number 2, November, p. 173 - 180.

Blalock, H., 1964: **Causal Inferences in Nonexperimental Research,** University of North Carolina Press, Chapel Hill, North Carolina.

Bonnette, B., 1980: "Interactive Enumeration District Determination System", **Proceedings of the Fifth Annual Workshop of the National Capital Geographic Information Processing Group,** D.R.F. Taylor (ed.), p. 110 - 112.

Bradley, D.R., 1981: "Computer Assisted Collection Maps for the 1981 Census of Canada", **Proceedings of the Seminar on the State and Utility of Computer-Assisted Map and Chart Production,** National Advisory Committee on Control Surveys and Mapping, Ottawa, Canada, November 17-19.

Brockwell, S., 1989: Internal Statistics Canada Memorandum, May 3.

Broome, F.R. and Witiuk, S.W., 1980: "Census Mapping by Computer", in **Contemporary Cartography,** Volume 1, p. 191 - 219, D.R.F. Taylor (ed.), John Wiley & Sons, Toronto.

Bunge, W., 1962: "Theoretical Geography", **Lund Studies in Geography, Series C, General and Mathematical Geography,** Volume 1, University of Lund, Sweden.

Cameron, J.M., and Norcliffe, G., 1984: "Science and Politics in Electoral Districting: The Ward System in the Town of Vaughan, Ontario", **Ontario Geography,** No. 24, p. 63 - 80.

Cliff, A.D., and Haggett, P., 1970: "On the efficiency of alternative aggregations in region-building problems", 1970, **Environmental Planning,** Volume 2, p. 285 - 294.

Cliff, A.D., and Ord, J.K., 1975: "Model building and the analysis of spatial patterns in human geography", **Journal of the Royal Statistical Society**, Volume B, Number 37, 1975, p. 297 - 328.

Deichsel G., and Tubingen, D., 1980: "Random walk clustering in large data sets", **Compstat Proceedings**, International Association for Statistical Computing, p. 454 - 459.

Duffield, B.S., and Coppock, J.T., 1975: "The delineation of recreational landscapes: the role of a computer-based information system", **Transactions, Institute of British Geographers**, Number 66, p. 141 - 148.

Eilon, S.; 1972: "Goals and Constraints in Decision Making", **Operations Research Quarterly**, Volume 23, Number 1, p. 3 - 15.

Eilon, S., Watson-Gandy, C.D.T., and Christofides, N., 1971: Distribution Management: **Mathematical Modelling and Practical Analysis**, Griffen, London

Fellegi, I.P. and Weldon, J.I., 1967: "Computer Methods for Geographical Coding and Retrieval of Data in the Dominion Bureau of Statistics, Canada", **Proceedings of the Social Statistics Section, American Statistical Association**, Washington, p. 53 - 57.

Gaile, G.L., 1984: "Measures of Spatial Equality", **Spatial Statistics and Models**, G.L. Gaile and C.J. Willmott (eds.), D. Reidel Publishing Company, p. 223 - 233.

Garey, M.R., and Johnson, D.S., 1978: **Computers and Intractability - A Guide to the Theory of NP-Completeness**, W. H. Freeman and Company, San Francisco.

Garfinkel, R.S., and Nemhauser, G.L., 1970: "Optimal Political Districting by Implicit Enumeration Techniques", **Management Science**. Number 16, p. 495 - 508.

Gelinas, R. and van Wyngaarden, R., 1988: "Development of 'Environomic Units': A Hybrid Spatial Framework Designed to Integrate Environmental and Socioeconomic Data for Land Modelling", **Perspectives on Land Modelling Workshop Proceedings**, R. Gelinas, D. Bond, and B. Smit (eds.), Polyscience Pub. Inc., Montreal, pp. 10

Ghosh, A., and Rushton, G., 1987: **Spatial Analysis and Location/Allocation Models**, Van Nostrand- Reinhold, pp. 373.

Glick, B., 1990: Personal correspondence on spatial decision support systems for establishing telephone and retail marketing districts.

Godlund, S., 1961: "Population, regional hospitals, transport facilities and regions: planning the location of regional hospitals in Sweden", **Lund Studies in Geography**. University of Lund, Lund, Sweden, Volume B, Human Geography, Number 21.

Goehlert, R., 1981: "Reapportionment and Redistricting: A Selected Bibliography", **Public Administration Series: Bibliography**, Vance Bibliographies, Post Office Box 229, Monticello, Illinois.

Goodchild, M.F., and Hosage, C.M., 1983: "On Enumerating All Feasible Solutions to Polygon Aggregation Problems", **Proceedings of the 14th Annual Pittsburgh Conference on Modelling and Simulation**, p. 591 - 595.

Haggett, P., Cliff, A.D., and Frey, A., 1977: **Locational Methods**, Edward Arnold Ltd., London.

Hannan, M.T., 1971: **Aggregation and Disaggregation**, D.C. Heath, Lexington, Mass.

Harvey, D.W., 1969: **Explanation in Geography**. London.

Helbig, R.E., Orr, P.K., and Roediger, R.R., 1972: "Political Redistricting by Computer", **Communications of the ACM**. Volume 15, Number 8, p. 735 - 741.

Heller, I., 1960: "Some numerical experiments for an M x J flow shop and its decision - theoretical aspects", **Operations Research**, 8.

Hodgart, R.L., 1978: "Optimizing access to public services", **Progress in Human Geography**, Volume 2, Number 1, p. 17 - 48.

Hodgart, R.L., 1985: "Developments in Location/Allocation Modelling", **Geographia Polonica**, Volume 51, p. 265 - 274.

Honey, R., and Kohler, J.A., 1978: **Distance Effects of School District Reorganization**, Final Project Report No. 16, Institute of Urban and Regional Research, University of Iowa.

Kajanoja, J., 1975: "The Problem of Aggregation in Locational Models", **Dynamics of Urban Space**, A. Karlqvist, L. Lundqvist and F. Snickars (eds.), Saxon House/Lexington Books, 1975. p. 327 - 340.

Kimerling, A.J., Morrison, J.L., and Yatabe, P.M., 1973: "Compactness in Legislative Redistricting: A Comparative Study of Several Shape Measures", **Proceedings of the 33rd Annual Meeting of the ACSM**, American Congress of Surveying and Mapping, Washington, D.C., USA, March 11 - 16, p. 275 - 289.

Lea, A.C., 1973: "Location-Allocation Systems: An Annotated Bibliography", **Discussion Paper Number 13**, Department of Geography, University of Toronto, pp. 272.

Lee, D.R. and Sallee, G.T., 1970: "A method of measuring shape", **Geographical Review**. Volume 60, p. 555 - 563.

Makarovic, B., 1979: "From Progressive To Composite Sampling For Digital Terrain Models", **Geoprocessing**, Volume 1, p. 145 - 166.

Mandelbrot, B., 1977: **Fractals: Form, Chance, and Dimension**, W.H. Freeman and Company, New York.

Mandelbrot, B., 1983: **The Fractal Geometry of Nature**, W.H. Freeman and Company, New York.

Marlin, P.G., 1981: "Application of the Transportation Model to a Large Scale 'Districting' Problem", **Computing and Operations Research**, Volume 8, p. 83 - 96.

Marx, R., 1985: Personal correspondence, March 25, 1985.

Marx, R., 1987: Personal correspondence, November 2, 1987.

Massam, B.H., 1972: **The Spatial Structure of Administrative Systems**, Association of American Geographers, Resource Paper Number 12, Washington, pp. 37.

Massam, B.H., 1975: **Location and Space in Social Administration**, Edward Arnold, Ltd., London.

Massam, B.H., and Burghardt, A.F., 1968: "The administrative subdivision of Southern Ontario: an attempt at evaluation", **Canadian Geographer**, Volume 12, Number 3, p. 125 - 134.

Massam, B.H., and Goodchild, M.F., 1971: "Temporal Trends in the Spatial Organization of a Service Agency", **Canadian Geographer**, Volume 15, Number 3, p. 193 - 206.

Masser, I., and Brown, P.J.B., 1978: **Spatial Representation and Spatial Interaction**, Leiden, Martinus Nijhoff.

Mather, P.M., 1976: **Computational Methods of Multivariate Analysis in Physical Geography**, John Wiley, Chichester, Sussex.

Medvedkov, Y.V., 1966: "The Concept of Entropy in Settlement Pattern Analysis", **Papers of the XVIII Vienna Congress of the Regional Science Association**, p. 165 - 168.

Moellering, H. and Tobler, W.R., 1972: "Geographical variances", **Geographical Analysis**. Volume 4, p. 34 - 50.

Moore, E.F., 1959: "The shortest path through a maze", **Annals of the Computation Laboratory of Harvard University**. Harvard University, Cambridge, Mass., Volume 30.

Morrill, R.L., 1973: "Ideal and reality in reapportionment", **Annals of the Association of American Geographers**. Association of American Geographers, USA, Volume 63, Number 4, p. 463 - 477.

Morrill, R.L., and Symons, J., 1977: "Efficiency and Equity Aspects of Optimum Location", **Geographical Analysis**, Volume IX, p. 215 -225.

Nicholson, R.L., 1982: "Redistricting Cities: The Case of Buffalo New York", **Ontario Geography**, Number 20, p. 65 - 82.

Openshaw, S., 1978: "An empirical study of some zone-design criteria, **Environment and Planning**, A 10, p. 781-794.

Openshaw, S., Sforzi, F., and Wymer, C., 1982: "The urban social structure of Firenze: an investigation using individual and aggregate census data", **Paper presented at the Twenty - Second Regional Science European Congress**, Groningen, The Netherlands, August 24 - 27, pp. 17.

Pollack, M., and Wiebenson, W., 1960: "Solutions of the shortest-route problem: a review", **Operations Research**, Volume 8, p. 224 - 230.

Reid-Green, K.S., 1989: "The History of Census Tabulation", **Scientific America**, February, p. 98 - 103.

Reock, E.C. Jr., 1961: "A note: Measuring compactness as a requirement of legislative apportionment.", **Midwest Journal of Political Science**, Volume 5, Number 1, p. 70 - 74.

Rhynsburger, D., 1973: "Analytic Delineation of Thiessen Polygons", **Geographical Analysis**, Number 5, p. 133 - 144.

Robinson, V.B., and Strahler, A.H., 1984: "Issues in Designing Geographic Information Systems Under Conditions of Inexactness", **Proceedings of the Tenth International Symposium on Remotely Sensed Data**, Purdue University, p. 198 - 204.

Sammons, R., 1976: "Zoning systems for spatial models", **Reading Geographical Papers**, Number 52, Department of Geography, University of Reading, Reading, England.

Sammons, R., 1978: "A simplistic approach to the redistricting problem", **Spatial Representation and Spatial Interaction**, I. Masser and P. Brown (eds.), 1978, p. 71 - 94.

Scott, A.J., 1968: **Combinatorial Processes, Geographic Space and Planning**, Department of Town Planning, University College London, Discussion Paper Series, Number 1.

Scott, A.J., 1971: **Combinatorial programming, spatial analysis and planning**. Methuen & Co. Ltd., London, England.

Seguin, G., 1978a: **Guidelines for the Delineation of 1981 Census Commissioner Districts**. Statistics Canada, Ottawa, Canada, 1978.

Seguin, G., 1978b: **Guidelines for the Delineation of 1981 Census Enumeration Areas**. Statistics Canada, Ottawa, Canada, 1978.

Slocum, T.A., Hanisch, R.J., and Thompson, D., 1984: "The Design and Application of a Cartographic Data Base for Redistricting", **Geo-Processing**, Volume 2, p. 151 - 176.

Statistics Canada, 1972: **GRDSR: Facts by small areas**. Statistics Canada, Ottawa, Canada.

Statistics Canada, 1980: **Report on Piece Rate Calculations for the 1981 Census**. Statistics Canada, Ottawa, Ontario, Canada.

Statistics Canada, 1981: **The Geography of the 1981 Census of Canada**. Statistics Canada, Ottawa, Ontario, Canada.

Statistics Canada, 1984: **Report on the Calculation of the 1981 Census Average Enumeration Time**. Statistics Canada, Ottawa, Ontario, Canada.

Statistics Canada, 1987: **1986 Census Dictionary**, Statistics Canada, Ottawa, Ontario, Canada.

Statistics Canada, 1989: **Summary Report on the Computer Assisted Districting Project (CADP) Prototype Volume Test (Laval)**. Statistics Canada, Ottawa, Ontario, Canada.

Steiner, D., 1965: "A multivariate statistical approach to climatic regionalization and classification", **Tijdschrift van het Koninklijk Nederlandsche Aardrijkskundig Genootschap**, Volume 15, p. 23 - 35.

Sutcliffe, C. and Board, J.L.G, 1986: "Designing secondary school catchment areas using goal programming", **Environmental and Planning A**, Volume 18, p. 661 - 675.

Tamminen, M., 1984; "Efficient Geometric Access To A Multirepresentation Geo-Database". **Geoprocessing**, Volume 2, p.177 - 196.

Taylor, D.R.F., 1980a: "Canadian and International Perspectives on Geocartographic Information Systems, **Proceedings of the Fifth Annual Workshop of the National Capital Geographic Information Processing Group**, Department of Geography, Carleton University, Ottawa.

Taylor, D.R.F., 1980b: **The Computer in Contemporary Cartography**. John Wiley & Sons, Toronto, pp. 252.

Thalmann, N., Claude, Y. Laporte, G. and Rousseau, J-M., 1982: "ELECT - An interactive graphical system for the automatic generation of electoral maps.", **Cartographica**, Volume 19, Number 1, p. 28 - 40.

Thompson, D., and Slocum, T., 1982: "A Geographic Information System For Political Redistricting in Maryland", Proceedings of the Applied Geography Conferences, Number 5, p. 73 - 87.

Tobler, W.R., 1973: "A Continuous Transformation Useful for Districting", **Annals of the New York Academy of Sciences**, Volume 219, p. 215 - 220.

Van Est, J., and De Vroege, F., 1983: "Accessibility and Segment Files: A Case Study of Districting Voting Areas", Paper Presented at the 10th Transportation Planning Research Colloquium, Zandroort, December 14-16, pp 19.

Weaver, J.B. and Hess, S.W., 1963: "A procedure for non-partisan districting: development of computer techniques", **Yale Law Journal**. Volume 73, p. 288 - 308.

Williams, I.N., 1976: "Optimistic theory validation from spatially grouped regression: theoretical aspects", **Transactions of the Martin Centre**, Number 1, p. 113 - 145.

Witiuk, S.W., and Brockwell, S., 1990: "Controlling a Spatial Decision Support System for Autodistricting Using Spatial Dispersion Indices", Paper for Presentation at SORSA '90, Freiburg, Germany, July.

Yeates, M.H., 1963: "Hinterland delimitation: a distance minimization approach", **Professional Geographer**, Volume 15, Number 6, p. 7 - 10.

APPENDIX A

NOTATION, FORMULAE AND ACRONYMS

A.1 NOTATION

- m is the desired number of districts (so is T);
- n is the number of building block units (so is N);
- G is a set containing the classification of n building blocks into m districts;
- A_k is that subset of G which contanins the indices of the building blocks allocated to the kth district;
- p_i is the population of the ith building block;
- b is the coefficient of the independent variable (i.e., ay=bx+c) and is an ordinary least squares estimate;
- x is the independent variable;
- y is the dependent variable;
- x_i is the value of x for the ith district;
- y_i is the value of y for the ith district;
- x_k is the mean value of x for the kth district;
- y_k is the mean value of y for the kth district;
- T is the target number of districts for the CTP (a priori or derived);
- V is the target value for each district in the CTP (a priori or derived);
- A is the surface area of the CTP;
- a is the expected mean area per district;
- a_i is the area of the ith building block;
- d_i is the deviation from the expected mean value (for area, block count, etc.);
- D is the sum of the absolute deviations d_i ;
- N is the total number of building block units in a CTP;
- ni is the number in each district; {change to n(j)}
- n is the expected mean number of blocks in each district;
- P is the target perimetre for each district in the CTP;
- L is the target length for the street network in each CTP:
- S is the target surface area for each CTP;
- p is the expected mean perimetre for each district;
- s is the expected mean surface area for each target;
- **D is the mean of the absolute deviations for a given CTP;**
- p_i is the actual perimetre for each block;

s_i is the actual surface area for each block;

H is the total number of dwellings in a CTP;

h_i is the actual number of dwellings in each building block;

A.2 FORMULAE

(1) $F(G) = \Sigma_k^m \mid \Sigma_{i \epsilon A(k)} p_i - \Sigma_i^n (p_i / m) \mid$

(2) $$F(G) = \frac{\Sigma_k^m \Sigma_{i \epsilon A(k)} (y_i - y_k)^2}{\Sigma_i^n (y_i - \Sigma_i^n y_i / n)^2} - \frac{\Sigma_k^m \Sigma_{i \epsilon A(k)} (x_i - x_k)2}{\Sigma_i^n (x_i - \Sigma_i^n x_i / n)^2}$$

(3) $d_i = a_i - a$; $i=1,..,T.)$

(4) $D = \Sigma\{|d_i|\}$ or $D = \Sigma\{|a_i - a|\}$; $i = 1,..,T$

(5) $V = N / T$

(6) $d_i = n_i - n$; for $i = 1,..,T$

(7) $D = \Sigma\{|d_i|\}$ or $D = \Sigma\{|n_i - n|\}$; $i = 1,..,T$

(8) $P = L / T$

(9) $S = A / T$

(10) $D = \Sigma\{|p_i - p|\}/T$; or $D = \Sigma\{|s_i - s|\}/T$; $i = 1,..,T$

(11) $CD = D/(2*L)$ or $CD = D/A$

(12) $D = \Sigma\{|d_i|\}/M$; $i=1,..,M$ and M is the target number of districts minus 1 to allow for the residual district).

(13) $r_i = c_i / a_i$

(14) $r = \Sigma\{r_i\}/N$ for $i=1,..,N$

(15) $SD = (\Sigma\{(r_i - r)**2.0\}/N)**0.5$

(16) $CV = SD / r$

(17) $SCV = CV/(N)**0.5$

(18) $FOM = \Sigma\{w_i * e_i * t_i\}$ for $i=1,..,$number of components,

(19) $VDC=1+\{a_i/A\}+\{1-h_i/H\}$

(20) $D = SUM\ \{d_i\}/M$; $i=1,..,M$ and where the 'deviations', d_i, are calculated as follows:

- o between the lower and upper bound (i.e., between 1400 and 1600 minutes), $d_i = 0$;
- o below the lower bound (i.e., 1400 minutes), d_i = (lower bound - PFOM value) which is set to zero for the residual (i.e., smallest valued) district; and
- o above the upper bound (i.e., above 1600 minutes), d_i = ((1 + (PFOM value - upper bound)/4.0)^2)*4.0.

A.3 ACRONYMS

AMF	- Area Master Files
ARA	- Address Register Areas
ARC/Info	- A commercial geographic information system
CA	- Census Agglomeration Area
CADP	- Computer Assisted Districting Package
CD	- Census Division
CMA	- Census Metropolitan Area
CSD	- Census Subdivision
CT	- Census Tract
CTP	- 'Census Tract Part'
DCL	- Digital Equipment Company Control Language
DIME	- Dual Independent Map Encoding (File/System)
DND	- Department of National Defense
EA	- Enumeration Area (Canada)
ED	- Enumeration District (USA, UK)
EDP	- Electronic Data Processing
FED	- Federal Electoral District
FORTRAN	- A programming language ("FORmula TRANslator")
GBF	- Geographic Base File
G12A	- A standard Census form recording the nature and relationship of the collection unit to the geographic framework
HH	- Census Households
NTS	- National Topographic System (Canada)
TIGER	- Topologically Integrated Geographic Encoding and Referencing
UA	- Urban Area
UNIVAC I	- First commercial computer
XAVE	- The mean X value for the Bounding Box
YAVE	- The mean Y value for the Bounding Box

APPENDIX B

BIBLIOGRAPHY

Atkinson, M.P., 1979: "Database Systems", **Journal of Documentation**. Volume 35, Number 1, p. 49 - 91.

Atkinson, M.P., 1980: **Data Management for Interactive Graphics**. Department of Computer Science, University of Edinburgh, Edinburgh, Scotland.

Batty, M., 1976: **Urban Modelling: Algorithms, Calibrations, and Predictions**. Cambridge University Press, Cambridge, England.

Beardwood, J.E. and Kirby, H.R., 1975: "Zone definition and the gravity model: the separability, excludability and compressibility properties", **Transportation Research**. Volume 9, p. 363 - 369.

Berry, B.J.L., 1967: "Grouping and regionalization: an approach to the problem using multivariate analysis", **Studies in Geography**. Northwestern University, Volume 13, p. 219 - 251.

Boots, B.N., 1973: "Some models of the random subdivision of space", Geografiska Annaler. B, 55, p. 34 - 48.

Boustedt, O., 1961: "Some National Approaches to Delimiting Urban Boundaries", **Urban Research Methods**, J. Gibbs (ed.), p. 41 - 47.

Boyce, R.R. and Clark, W.A.V., 1964: "The concept of shape in geography", **Geographical Review**. Volume 54, p. 561 - 572.

Cheung, T., 1974: "An interactive graphic display for region partioning by linear programming", **Communications of the ACM**, Volume 17, p. 513 - 516.

Chinnappa, K., 1987: **Report on Common Spatial Indices**, Unpublished Internal Report, Statistics Canada, pp. 41.

Chorley, R.J. and Haggett, P., 1967: **Models in Geography**, Methuen & Co, London, England.

Cohon, J.L. 1978: "Multiobjective Programming and Planning". **Mathematics in Science and Engineering**, Volume 140, Academic Press, New York.

Cooper, L., 1963: "Location - Allocation Problems", **Operations Research**, Volume 11, p. 331 - 343.

Cooper, L., 1964: "Heuristic methods for location-allocation problems", **SIAM Rev.** Volume 6, p. 37 - 53.

Cooper, L., 1972: "The Transportation Location Problem", Operations Research, Volume 20, p. 94 - 108.

Coppock, J.T. and Johnston, J.H., 1962: "Measurement in human geography", **Economic Geography**. Volume 38, Number 2, p. 130 - 137.

Corbett, J. and Farnsworth, G., 1971: "Theoretical basis of dual independent map encoding (DIME)", **Proceeding of the Ninth Annual Conference**. Urban and Regional Information Systems Association, New Orleans.

Corbett, J., 1979: **Topological Principles in Cartography**. US Bureau of the Census, Washington, D.C., USA.

Cox, K.R., 1965: "The Application of Linear Programming to Geographic Problems", **Tijdschrift voor Econ. en Soc. Geografie**, Nov/Dec, p. 228 - 236.

Craig, W.J., 1973: "Reapportionment and the computer", **Law and Computer Technology**. Volume 6, p. 50 - 56.

Craig, W.J., 1977: "Redistricting", **Annals of the American Association of Geographers**. American Association of Geographers, USA, Volume 67, p. 641 - 645.

Forrest, E., 1964: "Apportionment by computer", **The American Behavioral Scientist**. Volume 8, p. 34 - 35.

Forrest, E., 1965: "Electronic Reapportionment Mapping", **Data Processing Magazine**. July, p. 52 - 54.

Freeman, H. and Pieroni, G., 1979: **Proceedings of a NATO Advanced Study Institute, Maratea, Italy**. Academic Press, Inc., London, June 18-29, pp. 369.

French, S. 1986: **Decision Theory: An Introduction to the Mathematics of Rationality.** Mathematics and Its Application, John Wiley & Sons, New York.

Friedman, H.P. and Rubin, J., 1967: "On some invariant criteria for grouping data", **Journal of American Statisticians.** Volume 62, p. 1159 - 1178.

Gearhart, B.C. and Liittschwager, J.M., 1969: "Legislative Districting by Computer", **Behavioral Science.** Volume 14, p. 404 - 417.

Gibbs, J.P., 1961: **Urban Research Methods.** D. Van Nostrand Co. Ltd., Toronto.

Goodchild, M.F., 1979: "The Aggregation Problem in Location - Allocation", **Geographical Analysis**, Volume 11, Number 3, p. 240 - 255.

Goodchild, M.F., 1984: "ILACS: A Location Allocation Model for Retail Site Selection, **Journal of Retailing**, N.Y. University Institute of Retail Management, Volume 60, Number 1, p. 84 - 100.

Goodchild, M.F., 1979: "The effects of generalization in geographical data encoding", **Proceedings of a NATO Advanced Study Intitute, Maratea, Italy**, Academic Press, H. Freeman and G. Pieroni, (eds.), p. 191 - 205.

Goodchild, M.F., and Booth, P.J., 1980: "Location and Allocation of Recreation Facilities: Public Swimming Pools in London, Ontario", **Ontario Geography**, Number 15, p. 35 - 51.

Goodchild, M.F., and Massam, B.H. 1969: "Some Least-Cost Models of Administrative Systems in Southern Ontario", **Geographical Analysis**, Series B, Volume 52, Number 2, p. 86 - 94.

Haggett, P., 1965: **Locational Analysis in Human Geography.** London, UK.

Haggett, P., and Chorley, R.J.,1969: **Network Analysis in Geography.** London.

Haggett, P., 1975: **Geography: A Modern Synthesis.** New York, USA.

Hagood, M.J., 1943: "Statistical methods for delineation of regions applied to data on agriculture and population", **Social Forces.** Volume 21, p. 288 - 297.

Hale, M.Q., 1966: "Computer Methods of Districting", in **Reapportioning Legislatures - A Consideration of Criteria and Computers**, Chapter 9, p. 96 - 124, H.D. Hamilton (Ed.).

Hamilton, H.D., 1966: **Reapportioning Legislatures - A Consideration of Criteria and Computers**. Charles E. Merrill Books, Inc., Columbus, Ohio, USA.

Harvey, D.W., 1969: "Conceptual and measurement problems in the cognitive-behavioral approach to location theory", **Northwestern University, Studies in Geography**. Northwestern University, Volume 17, p. 35 - 67.

Hess, S.W., Weaver, J.B., Siegfelt, H.J., Whelan, J.N., and Zitlau, P.A., 1965: "Nonpartisan political redistricting by computer", **Operations Research**, Number 13, p. 998 - 1006.

James, G., Cliff, A.D., Haggett, P. and Ord, J.K., 1970: "Some discrete distributions for graphs with applications to regional transport networks", **Geografiska Annaler**. Volume B, Number 52, p. 14 -21.

Jensen, P.A., 1971: "Optimal network partitioning", **Operations Research**. Volume 19, p. 916 - 932.

Kaiser, H.F., 1966: "A objective method for establishing legislative districts", **Midwest Journal of Political Science**. Volume 10, 1966, p. 80 - 96.

Kansky, K.J., 1963: "Structure of transport networks: relationships between network geometry and regional characteristics.", **University of Chicago, Department of Geography, Research Papers**. University of Chicago, Department of Geography, Chicago, USA, Volume 84.

Karlqvist, A. and Lundqvist, L., 1972: "A contact model for spatial allocation", **Regional Studies**. Volume 6, p. 401 - 419.

Keane, M., 1975: "The size of the region-building problem", **Environment and Planning**. Volume A7, p. 575 -577.

King, L.J., 1969: "The analysis of spatial form and its relation to geographic theory", **Annals of the Association of American Geographers**. Volume 59, 1969, p. 573 - 595.

Knapp, E. M., 1979: "Spatial data integration", **Proceedings of a NATO Advanced Study Institute, Maratea, Italy**, H. Freeman and G. Pieroni (eds.), p. 47 - 61.

Kopec, R.J., 1963: "An alternative method for construction of Theissen polygons", **Professional Geographer**, Volume 15, Number 5, p. 24 - 26.

Langley, S.C., 1974: **Linear programming applications to school collecting districts**. University of Bristol, Department of Geography Dissertation, Bristol, UK.

Liles, W.C. and Nugent, E.D., 1981: "Data Management considerations for Large Cartographic Information Systems, **Proceedings of the ACSM**. The American Congress on Surveying and Mapping, Washington, D.C., USA, February 22-27, pp. 12.

Lin, S., 1965: "Computer solutions of the travelling salesman problem", **Bell System Technical Journal**. Volume 44, 1965, p. 2245 - 2269.

Lukermann, F., 1966: "Empirical expressions of nodality and hierarchy in a circulation manifold", **East Lakes Geographer**. Volume 2, p. 17 - 44.

Marazana, F.E., 1964: "On the Location of Supply Points to Minimize Transport Costs", **Operations Research Quarterly**, Pergamon Press, Volume 15, 1964, p. 261 - 270.

Massam, B.H., 1980: **Spatial Search: Applications to Planning Problems in the Public Sector**, Volume 23, pp. 295.

Masser, I., Batey, P.W., and Brown, P.J.B., 1975: "The design of zoning systems for interation models", **London Papers in Regional Science 5 : Regional Science - new concepts and old problems**. Volume 5, Cripps, E.L. (ed), London, England, p. 166 - 187.

Masser, I. and Brown, P.J.B., 1975: "Hierarchical aggregation procedures for interaction data", **Environment and Planning**. Volume A 7, p. 509 - 523.

Miehle, W., 1958: "Link-length minimization in networks", **Operations Research**. Volume 6, p. 232 - 243.

Mills, G., 1967: "The determination of local government electorial boundaries", **Operations Research Quarterly**. Volume 18, p. 243 - 255.

Molinero, C.M., 1988: "School in Southampton: A Quantitative Approach to School Location, Closure and Staffing", **Journal of the Operational Research Society**, Volume 39, Number 4, p. 339 - 350.

Morrill, R.L., 1973: "Ideal and reality in reapportionment", **Annals of the Association of American Geographers**. Association of American Geographers, USA, Volume 63, Number 4, p. 463 - 477.

Morrill, R.L., 1974: "Efficiency and Equity of Optimum Location Models", **Antipode**, Volume 6, Number 1, p. 41 - 46.

Morrill, R.L., 1976: "Redistricting Revisited", **Annals of the Association of American Geographers**, Volume 66, Number 4, p. 548 - 556.

Nagel, S.S., 1971: "The theory and practice of bipartisan constitutional computer-aided redistricting", **Proceedings of the 1971 Spring Joint Computer Conference**. AFIPS Press, New Jersey, USA, Volume 38, p. 137 - 142.

Openshaw, S., 1976: " An emprical study of some spatial interaction models", **Environment and Planning**. Volume A 8, p. 23 - 41.

Openshaw, S., 1977: "A geographical solution to scale and aggregation problems in region building", **Transactions of the Institute of British Geographers**. New Series 2, p. 459 - 472.

Openshaw, S., 1977: "Optimal zoning systems for spatial interaction models", **Environmental Planning**. Volume A 9, p. 169 - 184.

Openshaw, S., 1977: "Algorithm 3: A procedure to generate pseudo-random aggregations of N zones into M zones, where M is less than N", **Environment and Planning**. Volume A, p. 1423 - 1428.

Openshaw, S., 1980: "Monothetic divisive algorithms for classifying large data sets", **Proceedings in Computational Statistics, COMPSTAT 1980**, M. M. Barritt and D. Wishart, (eds.), Physica-Verlag, Wien, p. 419 - 425.

Paddison, R., 1976: "Spatial Bias and Redistricting in Proportional Representation Election Systems: A Case - Study of the Republic of Ireland", **Tijdschrift voor Economische en Sociale Geografie**. Volume 67, Number 4, p. 239 - 241.

Pearson, A.D., 1974: "Heuristic approaches to road network optimisation", **Engineering Optimisation**. Volume 1, p. 37 - 49.

Perry, J.M., 1981: "Shifting Boundaries: Redistricting Battles Now Are Being Waged With Computer Maps", **Wall Street Journal**. Volume CXCVIII, Number 46, N.Y. USA, September 3, p. 1 & p. 16.

Peucker, T.K., 1979: "Computer cartography and the structure of its algorithms", **World Cartography**. Volume 15, p. 71 - 76.

Ridley, T.H., 1969: "Reducing the travel time in a transport network", **Studies in Regional Science**, A.J. Scott, (ed.), p. 73 - 87.

Robinson, V.B., and Frank, A.U., 1987: "Expert Systems for Geographic Information Systems", **Photogrametric Engineering and Remote Sensing**, Volume 53, Number 10, October, p. 1435 - 1441.

Rushton, G., 1989: "Applications of Location Models", **Annals of Operations Research**, Volume 18, p. 25 - 42.

Russo, R.L., Oden, P.H., and Wolff, P.K., 1971: "A heuristic procedure for the partitioning and mapping of computer logic graphs", **IEEE Transactions on Computers**. Volume C20, p. 1455 - 1462.

Sack, R.D., 1972: "Geography, geometry and explanation", **Annals of the Association of American Geographers**. Volume 62, p. 61 - 78.

Schoepfle, O.B., and Church, R.L., 1989: "School Redistricting with the aid of Quantitative Models and G.I.S.", Presented at the International Geographical Union Conference on Geography and Public Administration, Washington, D.C., March 17 - 19, pp. 19.

Scott, A.J., 1969: **Studies in Regional Science**. Pion, London.

Scott, A.J., 1969: "On the optimal partitioning of spatially distributed point sets", **Studies in Regional Science**, A.J. Scott (ed.), p. 57 - 72.

Silva, R.C., "Reapportionment and redistricting", **Scientific American**. Volume 213, Number 5, p. 20 - 27.

Sutcliffe, C., Board, J. and Cheshire, P., 1984: "Goal Programming and Allocating Children to Secondary Schools in Reading", **Journal of the Operational Research Society**, Volume 35, Number 8, p. 719 - 730.

Taafe, E.J., 1970: **Geography**. Englewood Cliffs, N.J., USA.

Tamminen, M., 1981: **Management of spatially referenced data**. Laboratory of Information Processing Science, Helsinki University of Technology, Finland, pp. 102.

Taylor, P.J., 1971: "Distances within shapes: an introduction to a new family of finite frequency distributions", **Geografiska Annaler**. Volume 53 B, p. 40 - 53.

Taylor, P.J., 1973: "Some implications of the spatial organisation of elections", **Transactions of the Institute of British Geographers**. Volume 60, p. 121 - 136.

Taylor, P.J., 1975: **Distance decay models in spatial interaction**. Norwich, UK.

Tobler, W.R., "Computation of the correspondence of geographical patterns", **Regional Science Association, Papers**. Regional Science Association, Volume 15, p. 131 - 142.

Tobler, W.R., 1966: "Notes on the Analysis of Geographical Distributions", **Michigan Inter-University Community of Mathematical Geographers Discussion Papers**. Number 8, pp. 15.

Tobler, W.R., 1982: "The Quadratic Transportation Problem as a Model of Interregional Migration Patterns.", **Collaborative Paper series**. International Institute for Applied Systems Analysis, A-2361 Laxenburg, Austria, December, pp. 14.

Toregas, C. and ReVelle, C., 1971: "Optimal Location Under Time and Distance Constraints", **Papers of the Regional Science Association**, Volume 28, p. 133 - 143.

Tornqvist, G., Nordbeck, S., Rystedt, B., and Gould P., 1971: Multiple Location Analysis, **Lund Studies in Geography**. Series C. General, Mathematical and Regional Geography, No. 12.

Ward, J.H., 1963: "Hierarchical grouping to optimise an objective function", **Journal of American Statistical Association**. Volume 58, p. 236 - 244.

Warren, W.G., 1962: "Contributions to the study of spatial point processes", **University of North Carolina, Institute of Statistics, Mimeo Series**. University of North Carolina, Institute of Statistics, North Carolina, USA, Volume 337.

Waugh, T.C., 1974: **Geographic Data in a System Environment**. University of Edinburgh, Scotland.

Weaver, J.B., 1970: **Fair and Equal Districts: A How-To-Do-It Manual on Computer Use**, National Municipal League, pp. 111.

Wildgen, J.K., and Engstgrom, R.L., 1980: "Computer Graphics and Political Cartography: ASPEX of Gerrymandering", pp. 20. {incomplete}

Wolfgram, D.D., Dear, T.J., and Galbraith, C.S., 1987: **Expert Systems for the Technical Professional**, John Wiley & Sons, New York.

APPENDIX C

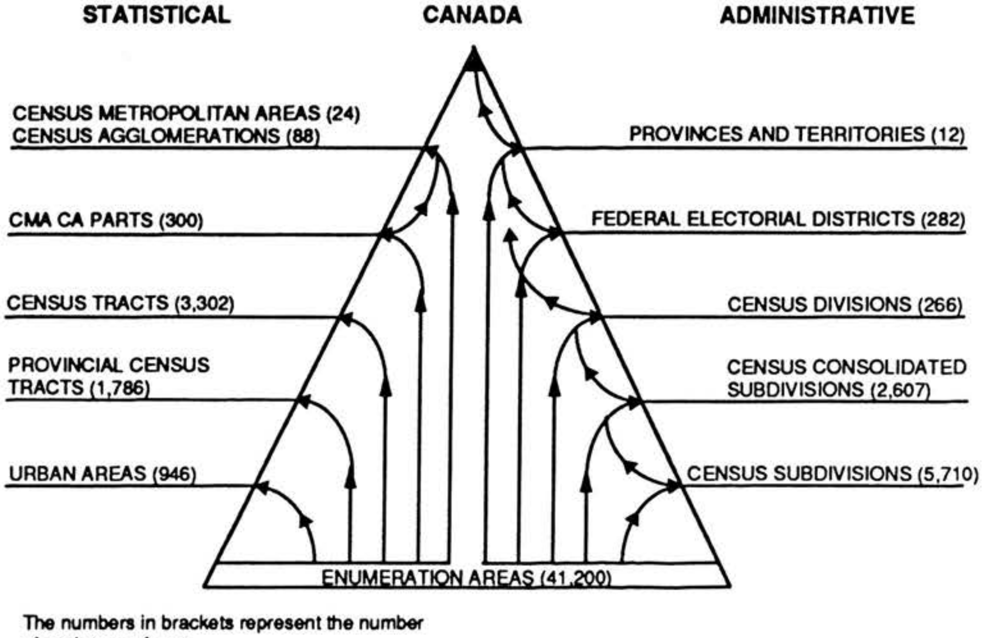
MANUAL DISTRICTING METHODS

C.1 INTRODUCTION

The methodology outlined here for delineation of enumeration areas is that employed by Statistics Canada for the taking of the Canadian Census. Other countries with similar geographical frameworks employ similar methods. Before each Census, field collection assignments - termed "enumeration areas" - are delineated by the Geography Division of Statistics Canada in accordance with specifications from the Survey Operations Division. In addition to permitting the efficient conduct of the field collection function, this delineation (or 'districting') must support the effective and accurate aggregation, retrieval and tabulation of the Census data according to pre-determined geographical and statistical classification schemes. Figure C.1 depicts the relationship between the enumeration area 'building blocks' and the standard Census geostatistical hierarchy. Descriptions of each of these spatial units are provided in the glossary.

FIGURE C.1

The 1981 Census Geographic Hierarchy



The numbers in brackets represent the number of each type of area.

* Approximate number

C.2 OVERVIEW

The following tasks comprise the major stages in the manual creation of enumeration areas:

A. Preparation of materials
   1. obtain and prepare map manuscripts
   2. obtain listings and counts (G12A, SRP, UR, Dwelling counts)
   3. transcription of geostatistical units to map manuscripts

B. Delineation Process
   1. identification of constraints
   2. reconciliation of map manuscripts and visitation records
   3. transcription of dwelling counts to map manuscripts
   4. transcription of EA counts to map manuscripts
   5. EA delineation (e.g., block grouping)

C. Verification and Correction
   1. analysis of districting results
   2. incorporation of revisions (e.g., municipal limit changes)
   3. modifications and/or improvements
   4. acceptance

D. Creation of Census Commissionaire Districts (CCDs)
   1. preliminary EA numbering
   2. grouping of EAs into CCDs
   3. preparation of CCDs summaries
   4. drafting of CCD limits

E. Evaluation and possible revision of the proposed districting in the regional field offices.

F. Distribution of field collection documents based on 'finalized' EAs

G. Post-Census revision of 'finalized' EAs changed during the collection process.

C.3 PROCEDURES

This section describes the guidelines and procedures that must be followed during each stage of the process overviewed in the previous section.

C.3.1 Preparation Of Materials

Over the years, Statistics Canada has generated a variety of map series' to support its Census-taking operations. These base maps range in scale from 1:400 to 1:500,000 and have been prepared from base maps produced by national, provincial, municipal and private mapping agencies. The typical sources can be summarized as follows:

Scale	Typical Source
Rural Canada	
1:500,000	Surveys and Mapping Branch, Energy, Mines and Resources Canada
1:250,000	Surveys and Mapping Branch, Energy, Mines and Resources Canada
1: 50,000	Surveys and Mapping Branch, Energy, Mines and Resources Canada
1:125,000	Provincial Mapping Agencies (e.g., B.C.)
Urban Canada	
1: 25,000	Military Charting Establishment, DND, and Surveys and Mapping Branch, Energy, Mines and Resources
1: 10,000	Military Charting Establishment, DND, and Surveys and Mapping Branch, Energy, Mines and Resources
1: 10,000	Engineering or Planning Departments of Municipal Governments
1: 4,800	Engineering or Planning Departments of Municipal Governments
1: 2,400	Engineering or Planning Departments of Municipal Governments
1: 1,000	Engineering or Planning Departments of Municipal Governments

The full range of pre- and post-Census map series' produced by Statistics Canada is described in the 1981 Census Product Guide and is summarized below:

Series G81-10: Provincial Maps

"The maps in this series show the boundaries of census divisions and census subdivisions. Census metropolitan areas are also shown. Except for Quebec and Ontario, an entire province is shown on one map sheet. In the case of Quebec and Ontario, seven and five maps, respectively, are required to show all of the CSDs for these two provinces. Map scales range from 1:250,000 in the most densely populated areas, to 1:4,000,000 in the Territories. An index map of all census divisions in Canada, at a scale of 1:5,000,000 is included in this series." [Statistics Canada, 1982, p. 128]

Series 81-12: Rural Enumeration Areas (EAs)

"The basic purpose of this series is to show the boundaries of EAs in the rural areas of Canada. The National Topographic System (NTS) series of maps, at scales ranging from 1:50,000 to 1:500,000 (and smaller scales for the Yukon and Northwest Territories), are used for this series of census maps. The G81-12 series also shows the boundaries of census divisions, census subdivisions, and federal electoral districts." [ibid]

Series G81-13: Census Tracts (CTs)/Enumeration Areas

"In this series, each map sheet covers one or more census tracts. Federal electoral district boundaries, enumeration area boundaries and street names, block numbers, and other physical features are identified. The G81-13 series is available for the 36 tracted centres; i.e., the 24 CMAs and 12 tracted CAs. The published CT maps for each CMA and CA provide an index to the individual CTs available in this series. The scales of maps in this series vary considerably." [ibid]

Series G81-13A: Census Tracts/Enumeration Areas

"The maps in this series are the same as those in the G81-13 series. However, for selected centres these maps are reduced to a standard size and are available in booklet form." [ibid]

Series G81-14: Urban Enumeration Areas

"The maps in this series show the boundaries of EAs located in the smaller urban centres containing four or more EAs. Most of the urban EAs not shown in the G81-13 series are included in the G81-14 series. The G81-14 series consists of place maps with enumeration area boundaries and street names. Map scales in this series vary." [ibid, p. 132]

Series G81-15: Census Divisions (CDs)/Census Subdivisions (CSDs) and Federal Electoral Districts (FEDs)/Enumeration Areas

"This series is provided by the Geographical Services Directorate, Surveys and Mapping Branch, Department of Energy, Mines and Resources. Complete coverage of the country consists of 41 map sets at 1:500,000, five regional maps sets at 1:2,000,000 and one map of the Northwest Territories and Yukon at 1:4,000,000. A set consists of a map showing census divisions and census subdivisions and a map showing federal electoral districts and census enumeration areas." [ibid, p. 132] Although this series is available for use in the creation of EAs for subsequent purposes, it is currently only used as a data dissemination instrument.

Series G81-18: Federal Electoral Districts/Enumeration Areas
"All federal electoral districts are covered by this series. Each map sheet covers one FED. All map sheets are produced at a scale large enough to sufficiently depict the EA boundaries. The urban FED map sheets show census tract as well as EA boundaries. The rural FED map sheets show the boundaries of census divisions, census subdivisions and enumeration areas." [ibid, p. 137]

Series G81-21: Provincial Census Tracts (PCTs)
"The purpose of this series is to show the boundaries of all PCTs. Those CMAs and CAs containing census tracts are also shown. National Topographic System maps are used as the bases for this series. Features such as water bodies, place names, road and railroads have been retained. Aside from provincial boundaries and the limits of census tracted CMAs and CAs, no other geostatistical areas (e.g., census subdivisions or enumeration areas) are shown on this series." [ibid, p. 137]

The full list is presented since post-Census map series' for one Census are often used as input to the enumeration area districting process of the subsequent Census.

In addition to the non-digital base map series' described above, an extensive data bank of cartographic information in digital form is also available for use by the manual districting process. Although they were not put to use in 1981, these files include:

Area Master Files
"The Area Master Files (AMFs) contain a logical representation of all city streets and other selected features (such as railroad tracks, rivers and municipal boundaries) in machine-readable form. The AMFs geographically reference every street, address range, block-face and centroid coordinate in the coverage area. An AMF is available for most centres with population of 50,000 or more." [ibid, p. 109]

Geography Tape File
"This file assigns to each 1981 census enumeration area (the basic geographic unit of census data collection) all higher geographic levels by codes and names. It includes all standard geographic codes, and also locates each enumeration area centroid according to Universal Transverse Mercator (UTM) coordinates and Lambert coordinates,

as well as by latitude and longitude. Final population counts for all enumeration areas are also presented." [ibid, p. 109]

Land Area and Population Tape Files

"Four tape files (one each for 1981 census divisions, census subdivisions, census tracts and provincial census tracts) present 1981 Census of Canada final population counts and land areas (in square kilometres), as well as the 1976 Census of Canada final population counts adjusted to the 1981 census geographic areas." [ibid, p. 109]

While some attempts were made during the 1981 Census to utilize this digital information as an aid to the manual districting process, little practical success was evident.

C.3.2 Delimitation Of The Geostatistical Boundary Hierarchy

Since geostatistical boundaries and, consequently, enumeration areas are likely (particularly in rural areas) to span several standardized map sheets, it is important to choose one of the geostatistical areas as the 'basic manuscript unit' for the districting process. All remaining geostatistical areas are then essentially 'cross-classified' by this basic work unit. In rural Canada, the basic work unit is the Federal Electoral District (FED) and sheets of the G-12 series are pieced together to portray entire FEDs. (Note: FEDs are re-apportioned every ten years based on successive decennial censuses, and therefore the composite manuscripts have a limited life-span.)

In urban Canada, the Census Tract (CT) is used as the basic work unit and sheets of the G-13 series are designed to comprise individual Census Tracts. Since Census Tracts are intended to provide a stable geographic framework for longitudinal studies, these manuscripts typically have a longer life-span than those for rural areas.

Working copies of the selected manuscripts are then prepared for the districting process:

1. Once the map sheets have been assembled into manuscripts of a manageable size for districting, measurement, field check and drafting operations, the census year, the NTS (National Topographic Survey) scale and index number are assigned to the manuscript to facilitate storage and retrieval.

2. Next, the current Federal Electoral District boundary is delimited and verified. All remaining geostatistical areas are then delimited and colour-coded according to the following scheme:

Unit		Colour
FED	- Federal Electoral District	Carmine
CD	- Census Division (Counties)	Black
CSD	- Census Sub-Division	
	- Incorporated city, town or village	Orange
	- Indian Reserves	Yellow
	- other municipalities	Yellow
	- 'Sub-Divisions' in Nfld., N.S., etc.	Ultramarine
	- Electoral areas in B.C.	Ultramarine
CT	- Census Tracts	True Green
PCT	- Provincial Census Tracts	True Green
UA	- Urban Areas	Pink
EA	- Enumeration Areas	Lavender

Finally, collective dwellings, military and penal establishments, are indicated:

Unit	Size	Colour
Collective Dwellings:		Burnt Ochre
- Hotels, Motels, Tourist Homes	200 beds	
- Lodgings - houses or school residences	150 beds	
- YM/YWCAs, Missions Hostels	200 beds	
- Work camps	150 beds	
- Religious institutions	150 beds	
- Orphanages and children's homes	75 beds	
- Nursing homes, Old age homes, and Chronic Care Institutions	75 beds	
- General Hospitals	75 beds	
- Psychiatric Hospitals	75 beds	
- Juvenile delinquent homes	75 beds	
- Corrective and Penal Institutions	75 beds	
- Hutterite colonies		
- Jails		
- Military camps		
- Other		

C.4 THE METHOD OF DISTRICTING

The process of delineating enumeration areas is essentially an acquired art. Within the guidelines and constraints listed above, the persons learn - through trial and error - undocumented, subjective strategies resulting in a final districting that, based on whatever information is available for the given territory, conforms to the mandatory constraints and guidelines.

As new information comes available, perhaps from a field check or during the actual field enumeration process, changes are made to the EA boundaries to ensure their continued conformance with the constraints and guidelines. Since no written procedures for either performing the actual districting or for evaluating independently generated results have been generated, the final results can only be judged subjectively as either 'acceptable' or 'unacceptable'.

C.4.1 Revision Of The Proposed Districting Based On Field Checks

Since, in the general case, information used in the districting process stems from the previous census which is typically 3 to 4 years out of date at the time it is used and since much of this information is only available at the level of the previous enumeration areas, it is necessary to conduct pre-census field checks in selected parts of the country. The purpose of these field checks is to confirm that the proposed districts conform to the stated criteria and guidelines. If, for example, major changes have occurred since the previous census, revisions to the district boundaries will be necessary. After the field check has been completed, the enumeration area boundaries are 'finalized' and a unique identification code (Province #/FED #/EA #) is assigned.

C.4.2 Distribution Of Field Collection Documents Based On 'Finalized EAs'

Once the boundary and identification code of an enumeration area has been finalized, the colour-coded boundary and code information on the map manuscript is manually transcribed to a 'fine drawing' base map for subsequent reproduction. Sufficient numbers of copies are then produced to ensure that each census representative is provided with a graphic description of his/her geographic area of responsibility. These individual enumeration area maps together with composite maps for groups of EAs, known as census commissionaire districts, are then provided to the Regional Offices who supervise the census collection operation.

C.4.3 Post-Census Revision Of 'finalized' Enumeration Areas That Were Changed During The Census Collection Process.

As mentioned earlier, it may be necessary to make changes to 'finalized' enumeration areas in the field during the census collection process. Since municipal boundaries are "frozen" effective January 1st of a given census year, the main cause of changes to enumeration area boundaries is a sudden increase or decrease in settlement densities. Also, the kind of change is typically restricted to splitting a given EA into two or more new EAs.

C.5 CONCLUSIONS

C.5.1 Feasibility

Creating enumeration areas that conform to the mandatory criteria and guidelines listed above is regularly demonstrated to be feasible by clerical means.

C.5.2 Procedures

No rigorous procedures have been committed to written form to remove the subjective and individualized nature of the EA districting process.

C.5.3 Evaluation Criteria

Evaluation criteria are limited to binary decisions which classify the results as 'acceptable' or 'unacceptable'. No attempt to quantitatively assess the relative merits of alternative districting patterns is evident.

C.5.4 Discrete Versus Continuous Functions

Workload targets are assigned based on a discrete model of the settlement pattern of the country. Five settlement densities and a seven-level discrete function related to collection methods provide the framework for assigning workload targets (see Tables 3.6a, 3.6b, 3.7a and 3.7b in Chapter 3).

APPENDIX D

DEFINITIONS

The definitions of geographic terms and census concepts are presented here in summary form to assist readers unfamiliar with the Canadian Census. They have been taken from the **1986 Census Dictionary** [Statistics Canada, 1987].

D.1 GEOGRAPHIC TERMS

D.1.1 Census Agglomeration (CA)

The general concept of a census agglomeration (CA) is one of a large urbanized core, together with adjacent urban and rural areas which have a high degree of economic and social integration with that core.

A CA is defined as the main labour market area of an urban area (the urbanized core) of at least 10,000 population, based on the previous census. Once a CA attains an urbanized core population of at least 100,000, based on the previous census, it becomes a census metropolitan area (CMA).

CAs are comprised of one or more census subdivisions (CSDs) which meet at least one of the following criteria:

1. the CSD falls completely or partly inside the urbanized core;

2. at least 50% of the employed labour force living in the CSD works in the urbanized core; or

3. at least 25% of the employed labour force working in the CSD lives in the urbanized core.

Exceptions to the above delineation criteria may occasionally be made in certain special situations.

It should be noted that CA boundaries may not conform precisely with the main labour market area, since CAs must respect CSD limits.

D.1.2 Census Division (CD)

This term applies to census divisions, counties, regional districts, regional municipalities, and five other types of geographic areas made up of groups of census subdivisions.

In Newfoundland, Manitoba, Saskatchewan and Alberta, provincial law does not provide for geographic areas which are intermediate between the census subdivision and the province. Therefore, census divisions have been created by Statistics Canada in cooperation with these provinces. In all other provinces, the different types of census divisions and their limits are established by provincial law.

D.1.3 Census Metropolitan Area (CMA)

The general concept of a census metropolitan area (CMA) is one of a very large urbanized core, together with adjacent urban and rural areas which have a high degree of economic and social integration with that core.

A CMA is defined as the main labour market area of an urban area (the urbanized core) of at least 100,000 population, based on the previous census. Once an area becomes a CMA, it is retained in the program even if its population subsequently declines.

Smaller labour market areas, centred on urbanized cores of at least 10,000 population, are included in the census agglomeration (CA) program.

CMAs are comprised of one or more census subdivisions (CSDs) which meet at least one of the following criteria:

1. the CSD falls completely or partly inside the urbanized core;

2. at least 50% of the employed labour force living in the CSD works in the urbanized core; or

3. at least 25% of the employed labour force working in the CSD lives in the urbanized core.

Exceptions to the above delineation criteria may occasionally be made in certain special situations.

It should be noted that CMA boundaries may not conform precisely with the main labour market area, since CMAs must respect CSD limits. CMAs may also differ from metropolitan areas designated by local authorities for planning or other purposes.

D.1.4 Census Subdivisions (CSD)

This term refers to municipalities, Indian reserves, Indian settlements or unorganized territories. In Newfoundland, Nova Scotia, and British Columbia, the term also describes geostatistical areas that have been created by Statistics Canada in co-operation with the provinces as equivalents for municipalities.

D.1.5 Census Subdivision Types

Census subdivisions are classified into various types, according to official designations adopted by provincial or federal authorities. With the exception of unorganized territories, Indian reserves and Indian settlements, hamlets in the Northwest Territories and settlements in the Yukon Territory, the type indicates the municipal status of a CSD. The following list indicates the most common CSD types sequenced by their abbreviations:

BOR	Borough
C	City - Cite
CM	County (Municipality)
COM	Community
CT	Canton (Municipalite de)
CU	Cantons unis (Municipalite de)
DM	District (Municipality)
HAM	Hamlet
ID	Improvement District
LGD	Local Government District
LOT	Township and Royalty
MD	Municipal District
NH	Northern Hamlet
NV	Northern Village
P	Paroisse (Municipalite de)
PAR	Parish
R	Indian Reserve - Reserve indienne
RM	Rural Municipality
RV	Resort Village
SA	Special Area
SCM	Subdivision of County Municipality
SD	Sans designation (Municipalite)
S-E	Indian Settlement - Etablissement indien
SET	Settlement
SRD	Subdivision of Regional District
SUN	Subdivision of Unorganized
SV	Summer village
T	Town
TP	Township
UNO	Unorganized - Non organise
V	Ville
VC	Ville Cri

VK Village Naskapi
VL Village
VN Village Nordique

D.1.6 Census Tract (CT)

Census tracts are small, permanent census geostatistical areas established in large urban communities with the help of local specialists interested in urban and social science research. Census tracts are reviewed and approved by Statistics Canada according to the following criteria:

1. the boundaries must follow permanent and easily recognized lines on the ground;

2. the population must be between 2,500 and 8,000 with a preferred average of 4,000 persons, except for census tracts in the central business district, major industrial zones, or in peripheral rural or urban areas that may have either a lower or higher population;

3. the area must be as homogeneous as possible in terms of economic status and social living conditions; and

4. the shape must be as compact as possible.

All census metropolitan areas and all census agglomerations with a census subdivision having a population of 50,000 or more at the previous census are eligible for a census tract program. Once an urban centre is added to the program, it is retained even if its population subsequently declines.

While census tract boundaries do not necessarily respect census subdivision boundaries, they do respect the boundaries of census metropolitan areas and census agglomerations and their constituent primary census metropolitan areas and primary census agglomerations.

D.1.7 Enumeration Area (EA)

This term refers to the area usually canvassed by one Census Representative. It is defined according to the following criteria:

1. Households - the number of households in an enumeration area generally varies between a maximum of 375 households in large urban areas to a minimum of 125 in rural areas;

2. Limits - an enumeration area, being the building block of all geostatistical areas, never

cuts across any geographic area recognized by the census.

Moreover, enumeration area boundaries are defined such that the Census Representative will be able to locate them with as little difficulty as possible, for example, streets, roads, railways, rivers and lakes. Enumeration areas are normally the smallest geographic unit for which census data are available.

D.1.8 Federal Electoral District (FED)

This term refers to any territorial unit entitled to return a member to serve in the House of Commons. There are 282 FEDs in Canada based on the 1976 Representation Order. These FEDs are used both to present data and to organize census-taking.

D.1.9 Indian Reserve

The term 'Indian Reserve' refers to land, the legal title to which is vested in her Majesty, that has been set apart for the use and benefit of an Indian band and that is subject to the terms of the Indian Act. Since it is generally excluded from local jurisdiction and is administered by the Federal Department of Indian and Northern Affairs Canada (INAC), it is classified as a census subdivision (CSD) by Statistics Canada.

D.1.10 Indian Settlement

Indian Settlements are places, identified by the Federal Department of Indian and Northern Affairs Canada (INAC) for statistical purposes only, where a self-contained group of at least 10 Indian people reside more or less permanently. Indian settlements are usually located on Crown lands under federal or provincial jurisdiction. They have not been set apart for the use and benefit of an Indian band as is the case with Indian reserves.

D.1.11 Province

Provinces are the major political division of Canada. From a statistical point of view, it is a basic unit for which data are tabulated and cross-classified. In census publications, provincial tables include the Yukon Territory and the Northwest Territories.

D.2 CENSUS CONCEPTS

D.2.1 Census Family

This term refers to a husband and a wife (with or without children who have never married, regardless of age), or a lone parent of any marital status, with one or more children who have never married, regardless of age, living in the same dwelling. For census purposes, persons living in a common-law type of arrangement are considered as now married, regardless of their legal marital status; they accordingly appear as a husband-wife family in most census family tables.

D.2.2 Dwelling, Collective

This term refers to a dwelling of a commercial, institutional or communal nature. It may be identified by a sign on the premises or by a Census Representative speaking with the person in charge or with a resident or a neighbour, etc. Included are rooming-or lodging-houses, hotels, motels, tourist homes, nursing homes, hospitals, staff residences, communal quarters of military camps, work camps, jails, missions, group homes, and so on. Collective dwellings may be occupied by usual residents or solely by foreign and/or temporary residents.

D.2.3 Dwelling, Occupied Private

This term refers to a private dwelling in which a person or group of persons is permanently residing on Census Day. Also included are private dwellings whose usual residents are temporarily absent on Census Day. Unless otherwise specified, all data in housing reports are for occupied private dwellings rather than unoccupied private dwellings or dwellings occupied solely by foreign and/or temporary residents.

D.2.4 Dwelling, Private

The term 'Private Dwelling' refers to a separate set of living quarters with a private entrance either from outside or from a common hall, lobby, vestibule or stairway inside the building. The entrance to the dwelling must be one which can be used without passing through the living quarters of someone else.

D.2.5 Dwelling, Private, Occupied By Foreign And/Or Temporary Residents

This term refers to a private dwelling occupied solely by foreign and/or temporary residents on Census Day. A temporary resident of a dwelling is a person who resides there on Census Day, but has a usual place of residence elsewhere in Canada. A foreign resident is a person whose usual place of residence is outside Canada.

D.2.6 Ethnic Origin

'Ethnic Origin' refers to the ethnic or cultural group(s) to which the respondent or the respondent's ancestors belong. Ethnic or cultural group refers to the 'roots' or ancestral origin of the population and should not be confused with citizenship or nationality.

It should be noted that prior to the 1981 Census, only the respondent's paternal ancestry was to be reported. If multiple ethnic origins were reported, only one origin was captured, resulting in one ethnic origin per respondent. In 1981, this restriction was removed, allowing for multiple ethnic origins. One write-in was provided on the 1981 questionnaire, in addition to the mark boxes. The 1986 Census questionnaire allows respondents to write in up to three ethnic origins not included in the mark boxes. This increases the number of multiple response possibilities.

The 1986 question was changed slightly from that asked in the 1981 Census. In 1981, respondents were asked, "To which ethnic or cultural group did you or your ancestors belong on first coming to this continent?". The phrase "on first coming to this continent" was removed from the 1986 question. The 1986 ethnic origin question was: "To which ethnic or cultural group(s) do you or did your ancestors belong?"

D.2.7 Household

The term 'household' refers to a person or group of persons (other than foreign residents) who occupy a dwelling and do not have a usual place of residence elsewhere in Canada. It usually consists of a family group with or without lodgers, employees, etc. However, it may consist of two or more families sharing a dwelling, a group of unrelated persons, or one person living alone. Household members who are temporarily absent on Census Day (e.g., temporary residents elsewhere) are considered as part of their usual household. For census purposes, every person is a member of one and only one household. Unless otherwise specified, all data in household reports are for private households only.

Households are classified into three groups: PRIVATE HOUSEHOLDS, COLLECTIVE

HOUSEHOLDS and HOUSEHOLDS OUTSIDE CANADA.

D.2.8 Household, Collective

This term refers to a person or group of persons who occupy a collective dwelling and do not have a usual place of residence elsewhere in Canada. Data for collective households with foreign and/or temporary residents only are not shown.

D.2.9 Household, Private

The term 'Private Household' refers to a person or group of persons (other than foreign residents) who occupy a private dwelling and do not have a usual place of residence elsewhere in Canada. The number of private households equals the number of occupied private dwellings.

D.2.10 Household Type

'Household Type' refers to the basic division of private households into family and non-family households. Family household refers to a household that contains at least one census family (e.g. persons living in the same dwelling who have a husband-wife or parent and never-married child relationship). One-family household refers to a single census family that occupies one private dwelling. The family may be that of the person responsible for household payments (primary family) or a family in which the person responsible for household payments is not a member (secondary family). A multiple-family household is one in which two or more census families occupy the same private dwelling. Additional persons may or may not be present in such a household. A non-family household refers to one person who lives alone in a private dwelling, or to a group of persons who occupy a private dwelling and do not constitute a census family.

D.2.11 Institutional Resident

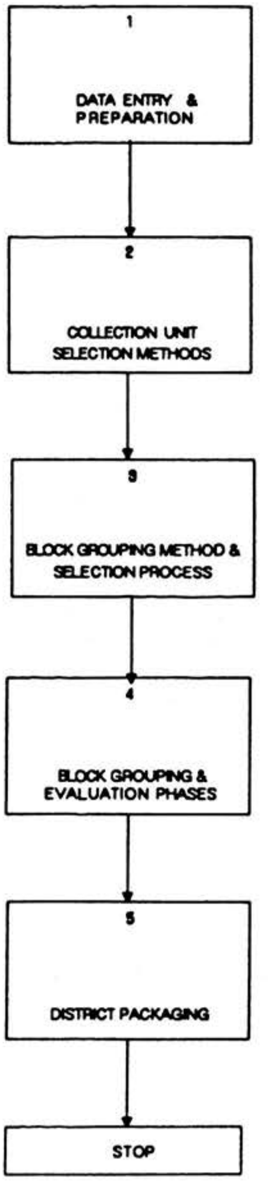
This term refers to a resident of an "institutional" collective dwelling, other than staff members and their families.

"Institutional" collective dwellings are orphanages and children's homes, special care homes and institutions for the elderly and chronically ill, hospitals, psychiatric institutions, treatment centres and institutions for the physically handicapped, correctional and penal institutions, young offenders facilities and jails.

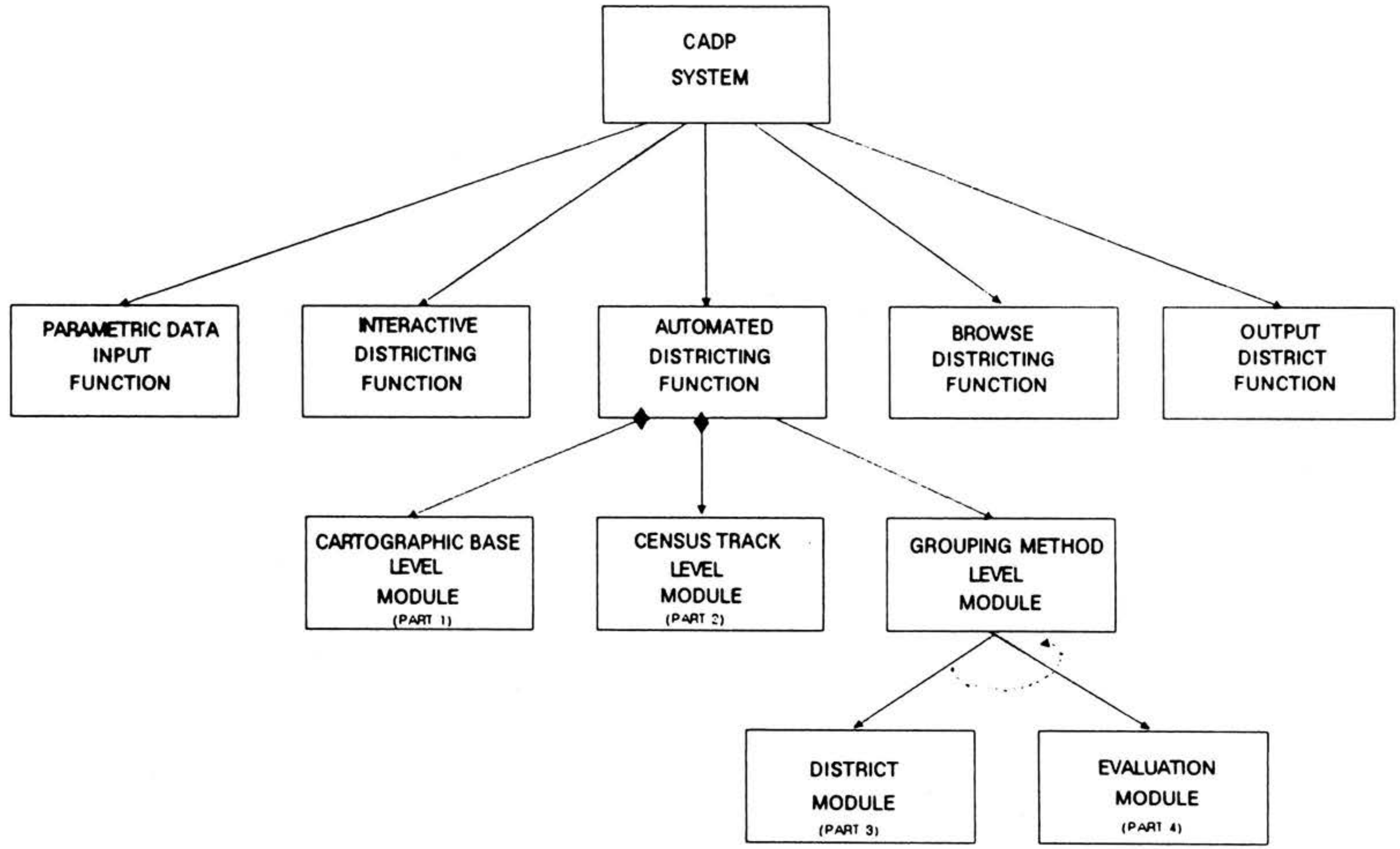
APPENDIX E

SYSTEM DOCUMENTATION

FLOW CHART OVERVIEW



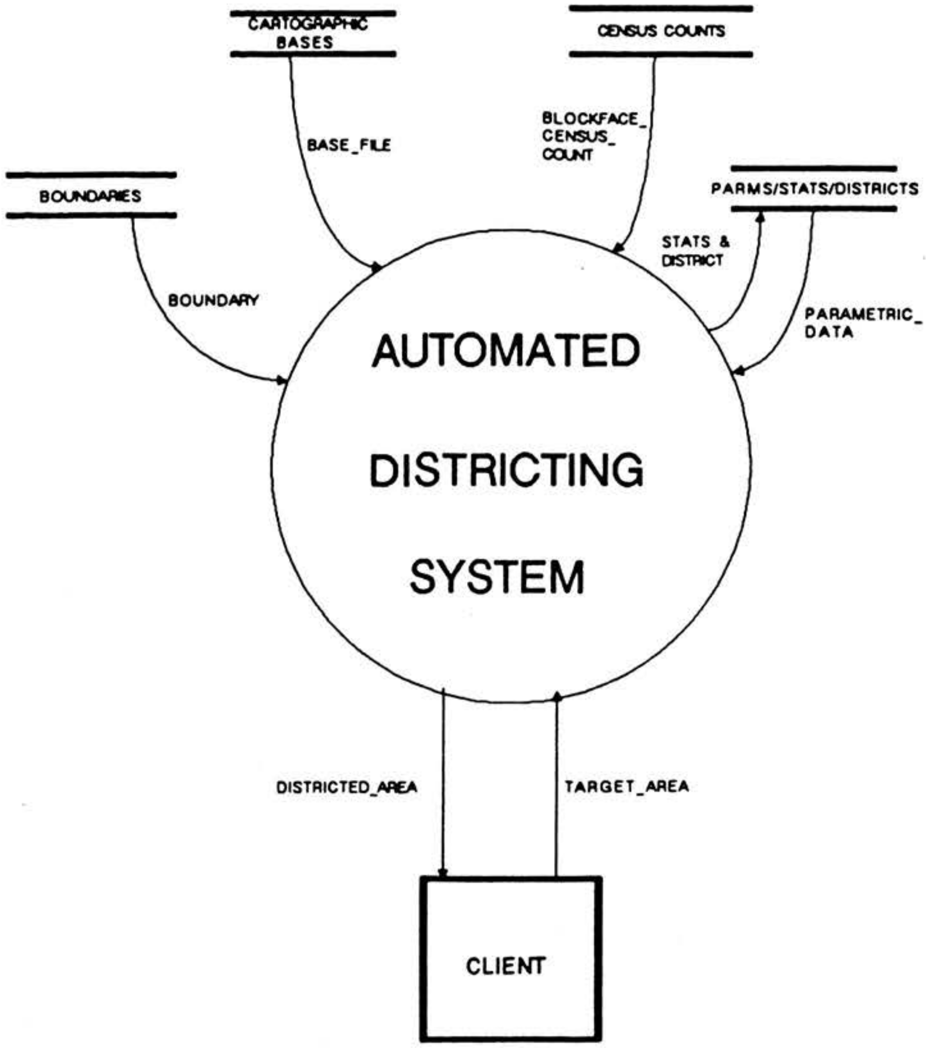
DISTRICTING SYSTEM (November 1987)



DISTRICTING SYSTEM

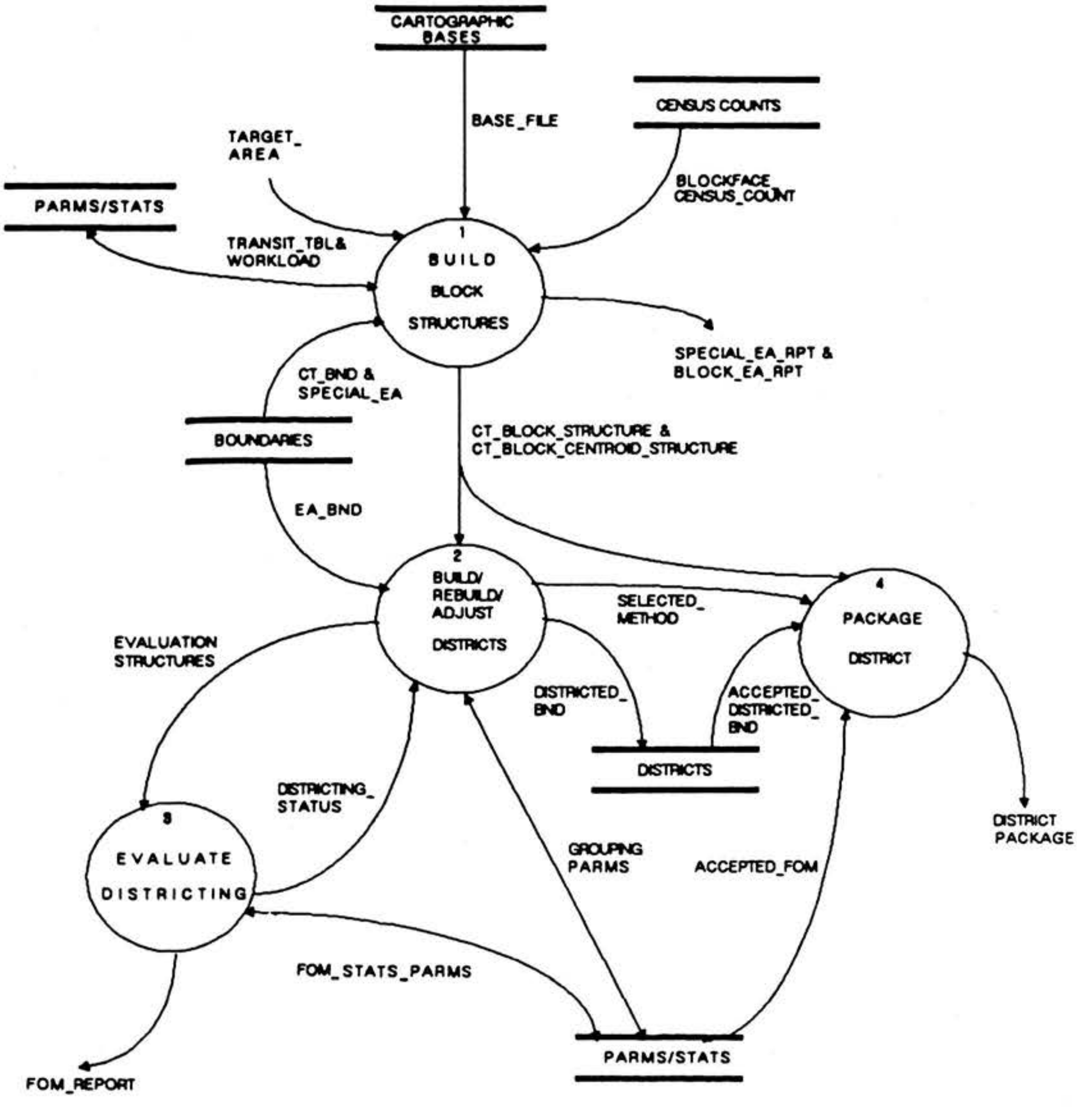
OVERVIEW

(November 1987)



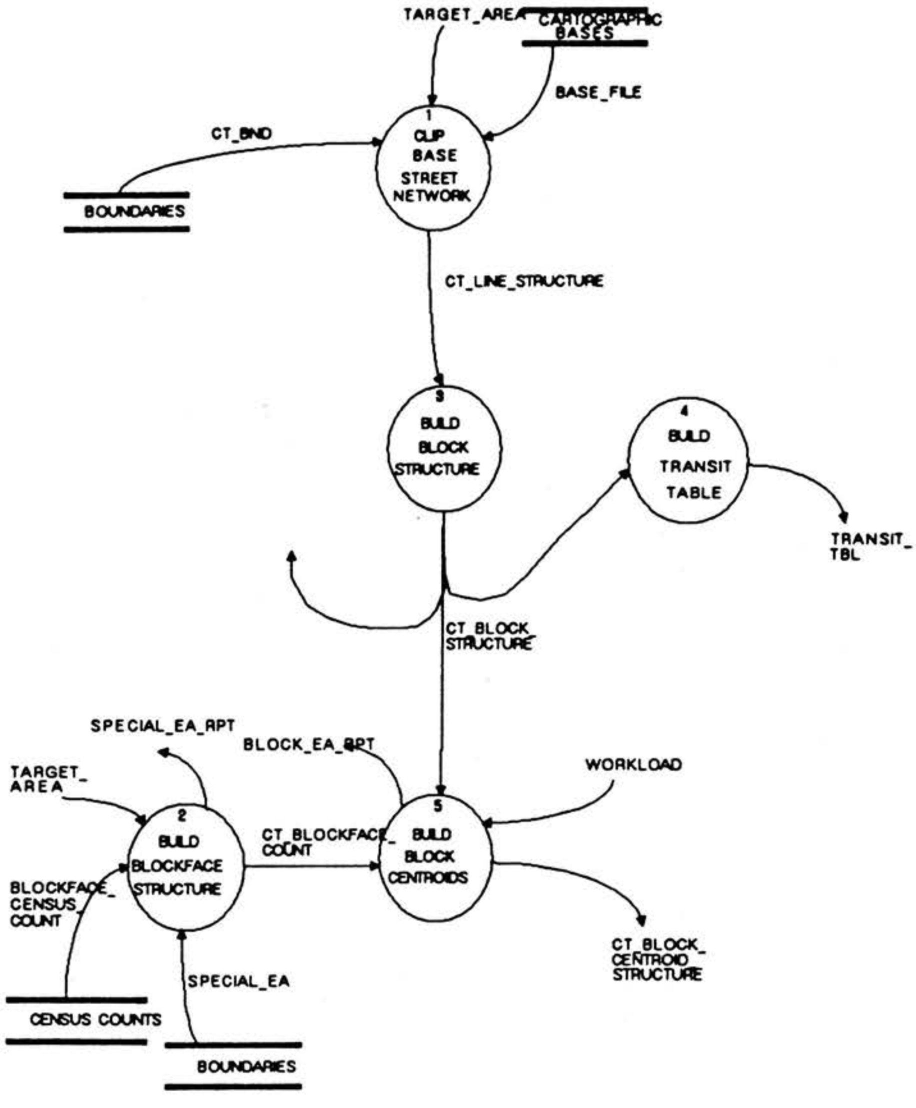
DISTRICTING SYSTEM

CONTEXT DIAGRAM
(November 1987)



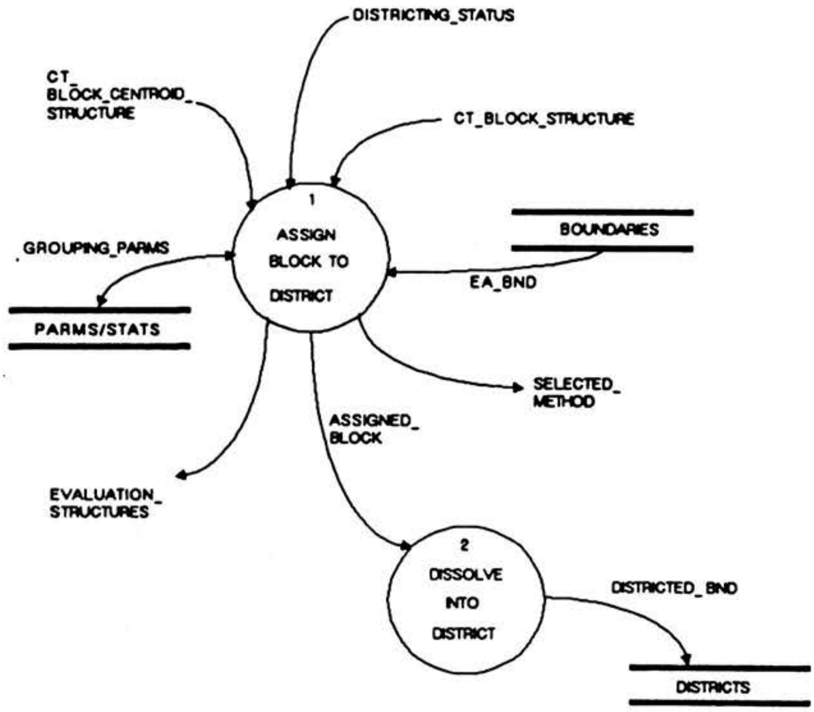
DISTRICTING SYSTEM

DIAGRAM: 0
(November 1987)



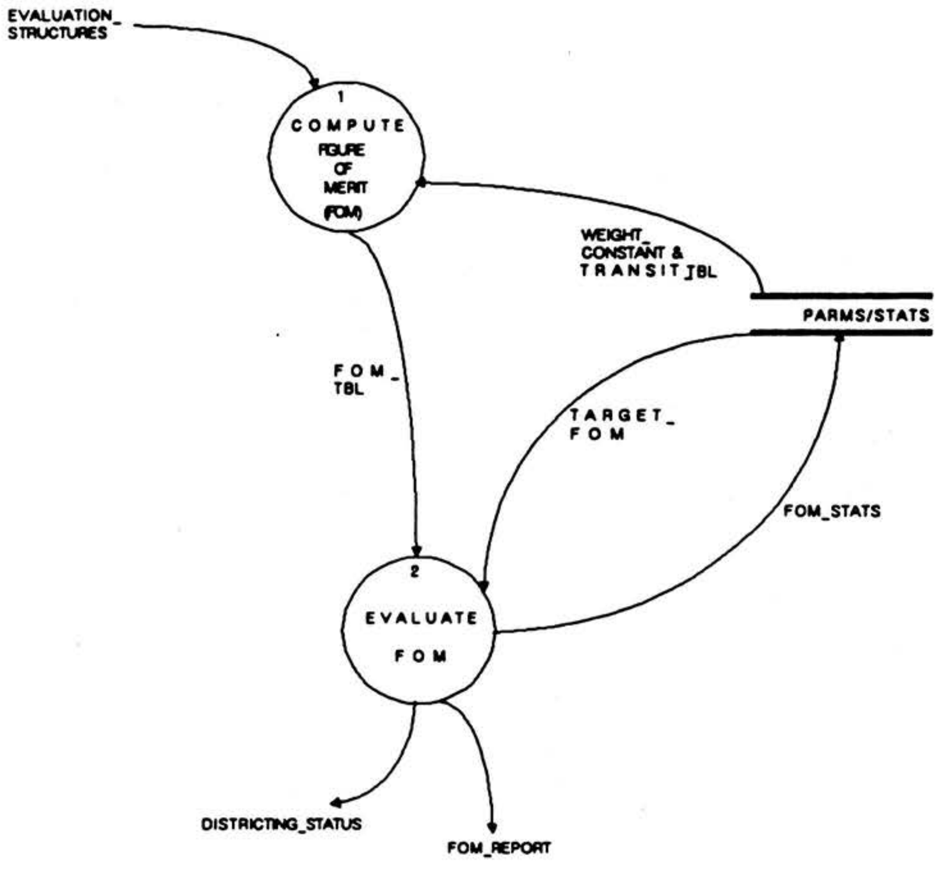
DISTRICTING SYSTEM

DIAGRAM: 1
(November 1987)



DISTRICTING SYSTEM

DIAGRAM: 2
(November 1987)

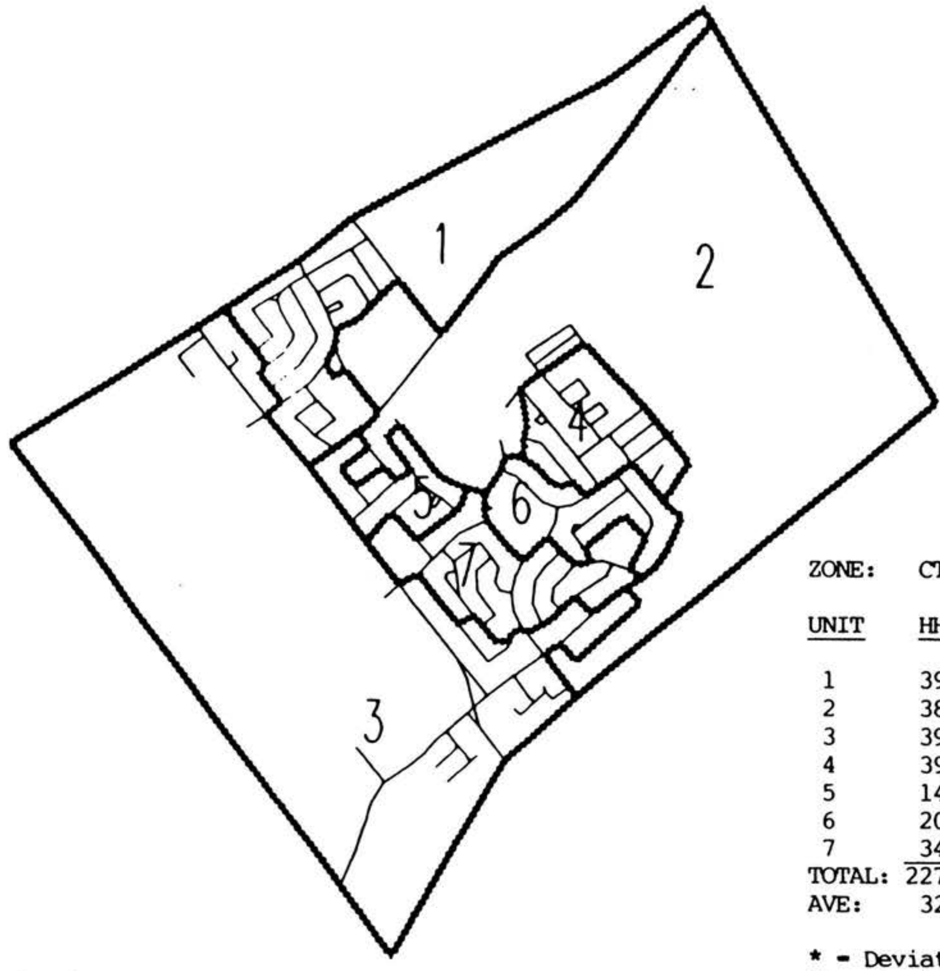


DISTRICTING SYSTEM

DIAGRAM: 3
(November 1987)

APPENDIX F

SAMPLE OUTPUTS



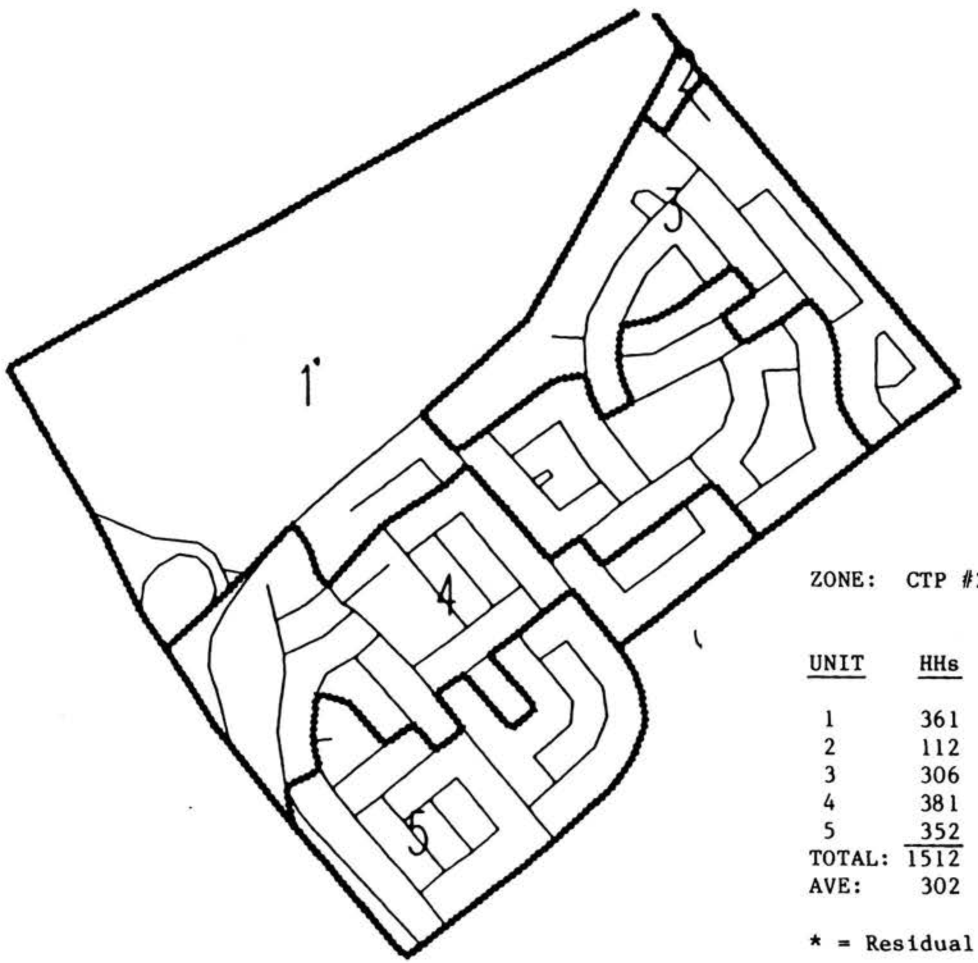
ZONE: CTP #24 TARGET - 7

UNIT	HHs	DEV*	PFOM	DEV*
1	393	0	1703	812
2	388	0	1682	324
3	398	0	1756	3025
4	395	0	1651	4
5	148	0**	640	0**
6	207	143	887	463
7	346	4	1487	0
TOTAL:	2275	147	9806	4628
AVE:	325	21	1401	661

* - Deviation PASS FAIL
** - Residual

SCALE AUTOMATIC
CTP NAME: 660 00 CTP CODE 019 718 - CTP 24
METHOD NINE UNIDIRECTIONAL
11/23/88 18 58.29
APPROACH FIVE. AGGREGATION BY DWELLING

ANNEX



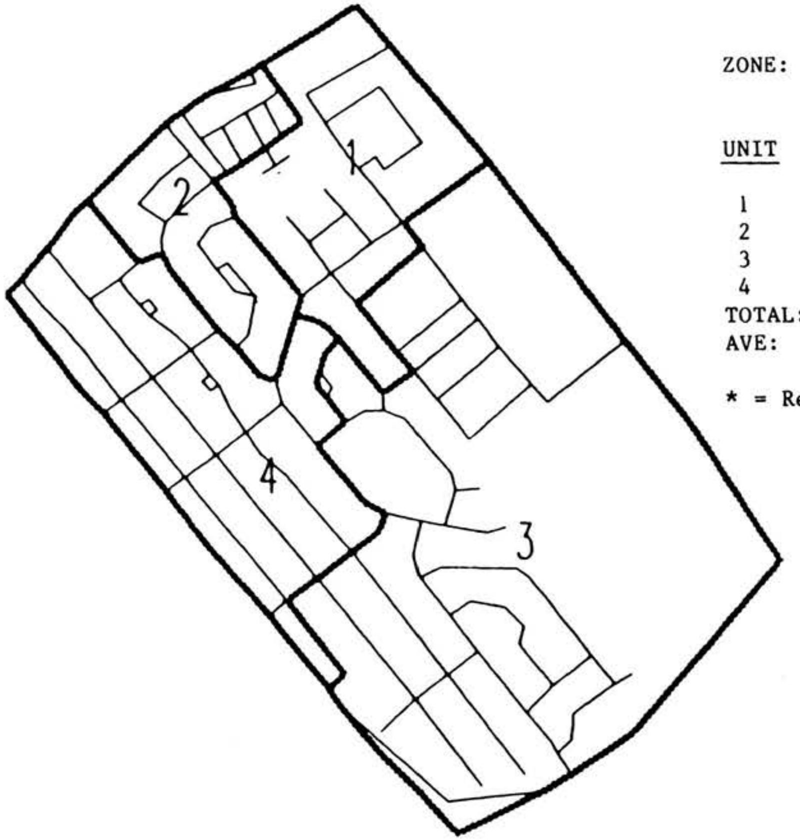
ZONE: CTP #27 TARGET = 5

UNIT	HHs	DEV	PFOM	DEV
1	361	0	1594	0
2	112	0*	440	0*
3	306	44	1326	24
4	381	0	1644	0
5	352	0	1513	0
TOTAL:	1512	44	6517	24
AVE:	302	9	1303	5

* = Residual PASS PASS

SCALE AUTOMATIC
CTP NAME: 630.02 CTP CODE: 0192080 - CTP: 27
METHOD NAME: UNIDIRECTIONAL GRID APPROACH FIVE: AGGREGATION BY DWELLING
11/19/88 03:12:18

CADP PROTOTYPE



ZONE: CTP #28 TARGET = 4

UNIT	HHs	DEV	PFOM	DEV
1	390	0	1606	0
2	260	0*	1098	0*
3	382	0	1713	1122
4	358	0	1553	0
TOTAL:	1390	0	5970	1122
AVE:	348	0	1492	280

* = Residual **PASS** **FAIL**

SCALE AUTOMATIC

CTP NAME 632 01 CTP CODE 0192091 - CTP 28

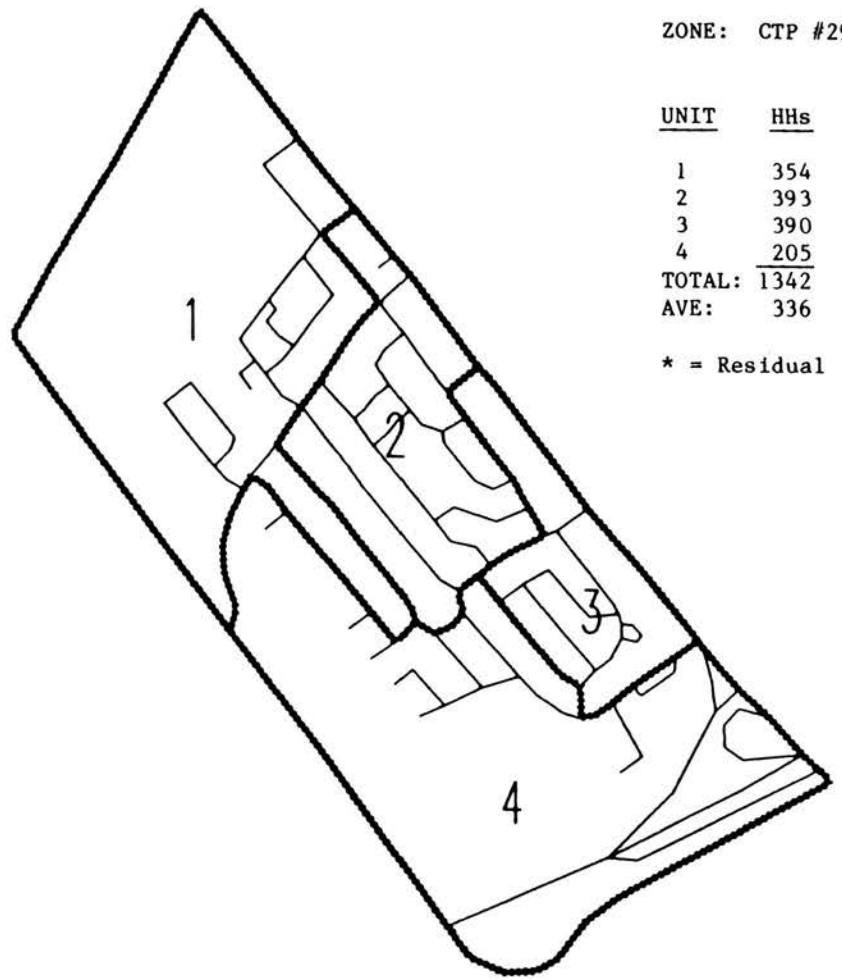
METHOD TWO: SECTION

APPROACH FIVE: AGGREGATION BY DWELLING

ZONE: CTP #29 TARGET = 4

UNIT	HHs	DEV	PFOM	DEV
1	354	0	1491	0
2	393	0	1652	5
3	390	0	1588	0
4	205	0*	895*	0
TOTAL:	1342	0	5626	5
AVE:	336	0	1406	1

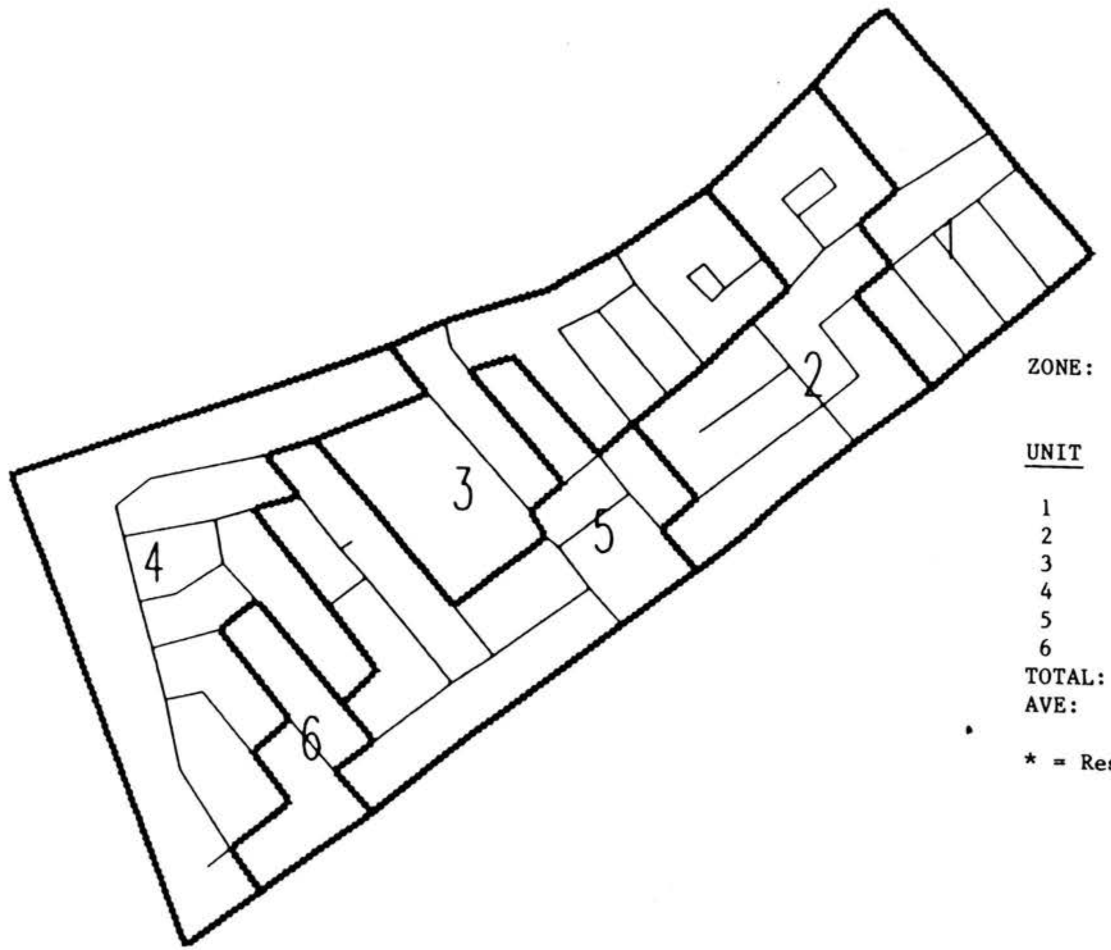
* = Residual PASS PASS



SCALE AUTOMATIC

CTP NAME 632.02 CTP CODE: 0192002 - CTP 29
METHOD NINE: UNIDIRECTIONAL GRID APPROACH FIVE: AGGREGATION BY DWELLING
11/23/88 20:29:15

ANNEX



ZONE: CTP #30 TARGET = 6

UNIT	HHs	DEV	PFOM	DEV
1	352	0	1423	0
2	232	0*	978	0*
3	391	0	1598	0
4	330	20	1371	0
5	387	0	1595	0
6	311	39	1230	120
TOTAL:	2003	59	8195	120
AVE:	334	10	1366	20

* = Residual PASS PASS

SCALE AUTOMATIC

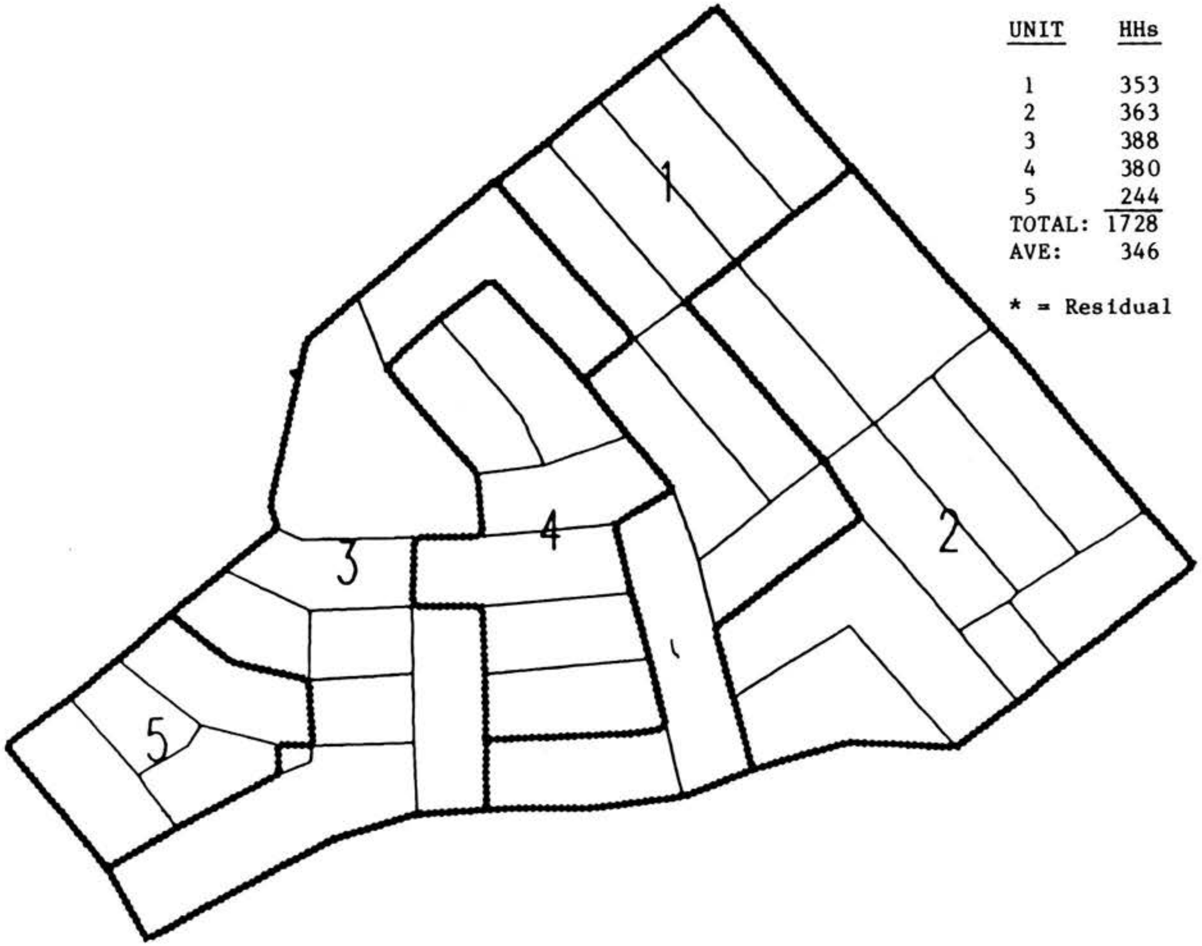
CTP NAME 638 01 CTP CODE 0362893 - CTP : 30
METHOD NINE: UNIDIRECTIONAL GRID APPROACH FIVE: AGGREGATION BY DWELLING
11/15/88 17:03:20

ANNEX

ZONE: CTP #32 TARGET = 5

UNIT	HHs	DEV	PFOM	DEV
1	353	0	1463	0
2	363	0	1505	0
3	388	0	1583	0
4	380	0	1530	0
5	244	0*	978	0*
TOTAL:	1728	0	7059	0
AVE:	346	0	1412	0

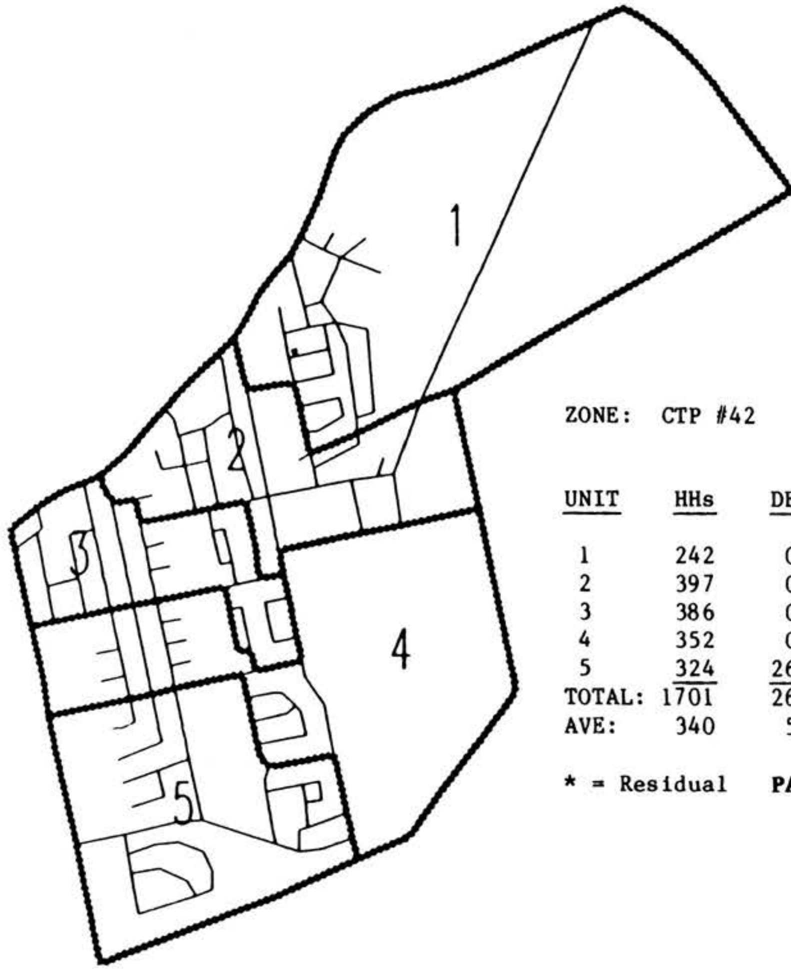
* = Residual **PASS** **PASS**



SCALE AUTOMATIC

CTP NAME: 630.03 CTP CODE: 0362095 - CTP : 32
METHOD NINE: UNIDIRECTIONAL GRID APPROACH FIVE: AGGREGATION BY DWELLING
11/19/88 22:07:24

CADP PROTOTYPE



ZONE: CTP #42 TARGET = 5

UNIT	HHs	DEV	PFOM	DEV
1	242	0*	1108	0*
2	397	0	1696	625
3	386	0	1641	0
4	352	0	1497	0
5	324	26	1425	0
TOTAL:	1701	26	7367	625
AVE:	340	5	1473	125

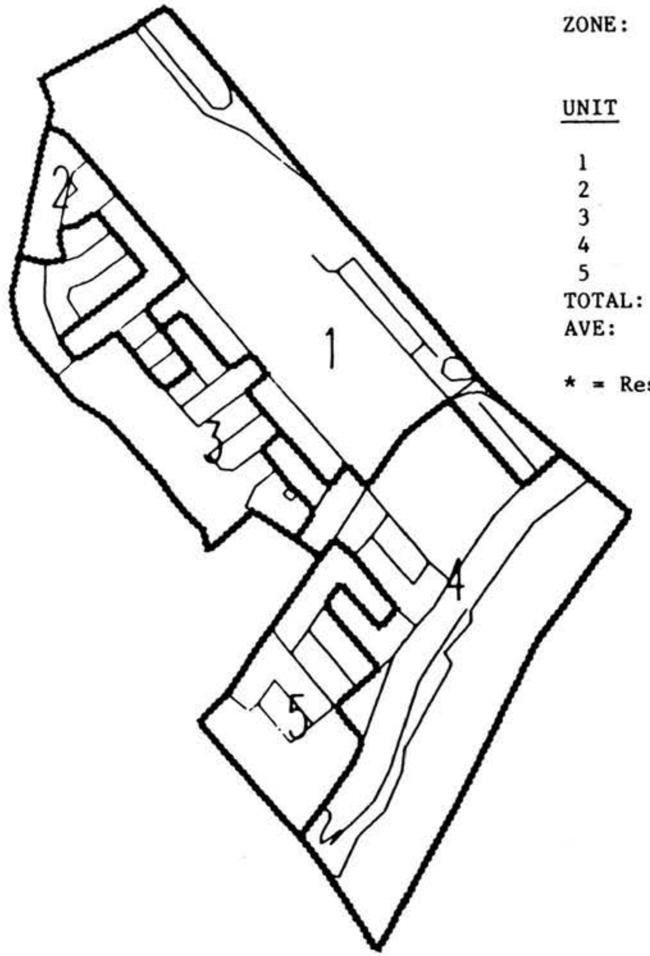
* = Residual **PASS** **PASS**

SCALE AUTOMATIC

CTP NAME: 659.02 CTP CODE: 0452905 - CTP : 42
METHOD NINE: UNIDIRECTIONAL GRID APPROACH FIVE: AGGREGATION BY DWELLING
11/15/88 10:34:08

ANNEX

CADP PG TYPE



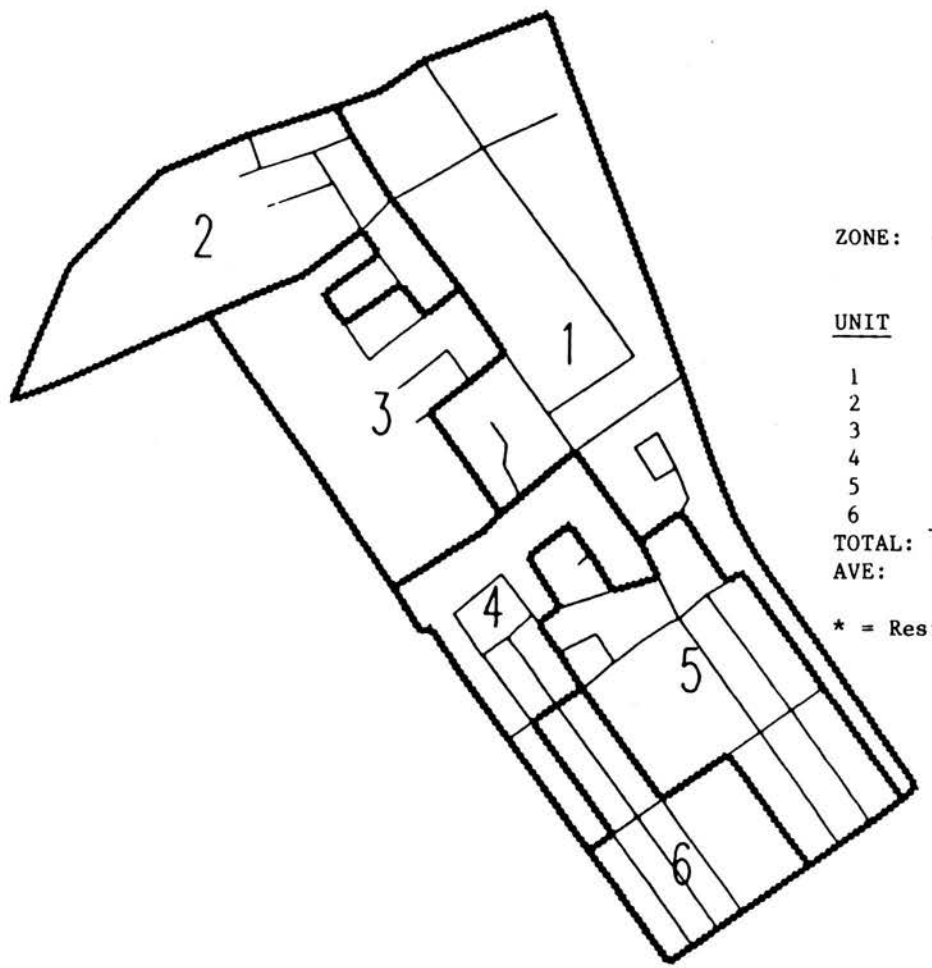
ZONE: CTP #43 TARGET = 5

UNIT	HHs	DEV	PFOM	DEV
1	394	0	1654	8
2	396	0	1599	0
3	361	0	1525	0
4	358	0	1533	0
5	208	0*	880	0*
TOTAL:	1717	0	7191	8
AVE:	343	0	1438	2

* = Residual **PASS** **PASS**

SCALE AUTOMATIC

CTP NAME 628.01 CTP CODE 0193152 - CTP : 43
METHOD TWO SECTOR APPROACH FIVE: AGGREGATION BY DWELLING
11/28/88 11:22:58 A.



ZONE: CTP #46 TARGET = 6

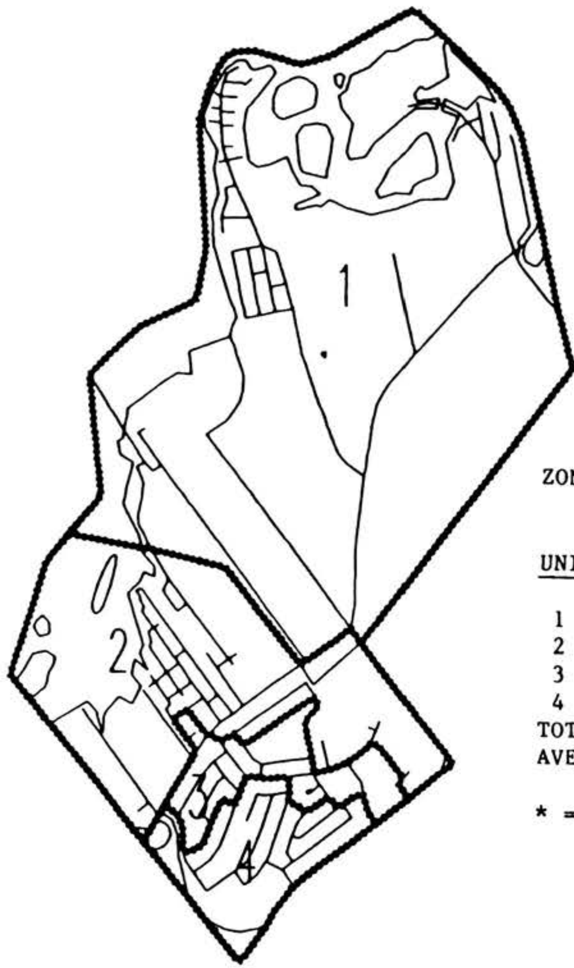
UNIT	HHs	DEV	PFOM	DEV
1	373	0	1558	0
2	388	0	1578	0
3	242	108	996	354
4	361	0	1466	0
5	391	0	1624	0
6	225	0*	939	0*
TOTAL:	1980	108	8161	354
AVE:	330	22	1360	59

* = Residual **PASS** **PASS**

SCALE: AUTOMATIC

CTP NAME 641 01 CTP CODE. 0363155 - CTP : 46
METHOD TWO, SECTOR APPROACH FIVE: AGGREGATION BY DWELLING
11/28/88 12.33:40

CADP PROTOTYPE



ZONE: CTP #52 TARGET = 4

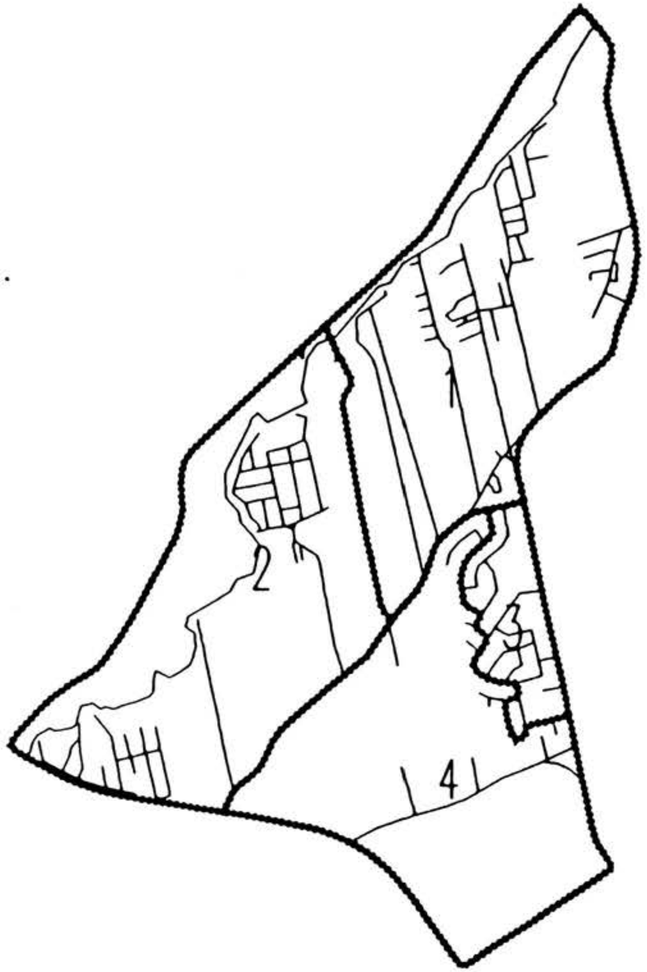
UNIT	HHs	DEV	PFOM	DEV
1	199	0*	1201	0*
2	399	0	1819	7482
3	327	23	1407	0
4	374	0	1595	0
TOTAL:	1299	23	6022	7482
AVE:	325	6	1505	1870

* = Residual **PASS** **FAIL**

SCALE AUTOMATIC

CTP NAME: 658.03 CTP CODE: 0453161 - CTP: 52
METHOD NINE: UNIDIRECTIONAL GRID APPROACH FIVE: AGGREGATION BY DWELLING
11/23/88 23:20:52

CADP PROTOTYPE



ZONE: CTP #53 TARGET = 4

UNIT	HHs	DEV	PFOM	DEV
1	376	0	1896	15625
2	315	35	1515	0
3	340	10	1445	0
4	244	0*	1108	0*
TOTAL:	1275	45	5964	15625
AVE:	319	11	1491	3906

* = Residual **PASS** **FAIL**

SCALE AUTOMATIC

CTP NAME: 661.01 CTP CODE: 0193162 - CTP: 53
METHOD TWO: SECTOR
11/15/88 10:45:29
APPROACH FIVE: AGGREGATION BY DWELLING

APPENDIX G

SUMMARY RESULTS OF THE VOLUME TEST FOR LAVAL

The tables presented in this appendix summarize the final results of over 629 test runs on the 63 Census Tract Parts of the Census Subdivision of Laval.

Each of the Census Tract Parts were evaluated based on dwelling count scores and on Partial Figure of Merit scores for:

1. actual 1981 districtings (63 runs);
2. dwelling-based districtings (236 runs); and
3. PFOM-based districting (204 runs)

for up to 11 alternative runs/steps for each approach for any given CTP.

The actual 1986 districts were subjected to the same automated evaluation process (126 runs). All evaluations (+/-) in the tables below are, therefore, relative to the actual 1986 districtings, which is used as the benchmark for: (a) the relative number of districts, (b) the PASS/FAIL status and (c) the magnitude of the variance. The actual PASS/FAIL status is determined in part by the target number of districts (last column of the table) and the variance threshold for the given approach.

The overall SCORE is determined by the first acceptable result in the sequence:

1. Previous (i.e., 1981) Districts
2. Dwelling-based results from CADP (called 'HH CADP') in Table G.1 or PFOM-based results (called 'FOM CADP') in Table G.2; and
3. PFOM-based results in Table G.1 or dwelling-based results in Table G.2.

Since the production system is intended to be a satisficing process rather than an optimizing process, it was deemed appropriate to rate the system's performance based on the first acceptable result rather than the best result of those tried.

The number of STEPS is determined from the following sequence:

0. Previous (1981) Districts
1. Method 9,
2. Method 9 with the ANNEX process,
3. Method 2,
4. Method 2 with the ANNEX process,
5. Method 6,
6. Method 6 with the ANNEX process,
7. REASSIGN based on the result of Method 9,
8. REASSIGN based on the result of Method 2,
9. REASSIGN based on the result of Method 6,
10. Exemption based on human intervention, and
11. Adjustment based on human intervention.

In all cases the process halts after an acceptable solution is obtained. The number of steps for the (a) dwelling-based and (b) PFOM-based CADP processes are provided as the first and second column, respectively, under the heading 'STEPS'. If the previous districts are re-used, the number of districting 'steps' is reported as zero (0).

Three factors are taken into consideration in the (+/-) assessments of the relative merits of the actual, manually-produced collection units and the manually-generated units from the previous Census or the two sets of autodistricting results as shown in Tables G.1 and G.4:

- o the relative number of districts generated (fewer are better);
- o the relative pass/fail status (pass is better); and
- o the relative magnitude of the variance values (smaller is better).

One '+' or '-' is given (up to a maximum of three) for each relative advantage and, if possible, a bonus is given for each 'extra' district that is saved.

Thus, comparative scores of 4F1.0 and 3P0.9 would result in a value of (+ + +) since 3 is less than 4, P is better than F and 0.9 is less than 1.0. Similarly, comparing 7P0.0 and 6P40.0 would give a value of (+ -) since 6 is less than 7 while 40.0 is greater than 0.0 and P = P.

TABLE G.1 METHOD COMPARISONS BASED ON DWELLING COUNT SCORES

CTP#	1986 EAs		1981 EAs		+/-	HH CADP		+/-	FOM CADP		+/-	SCORE	STEPS		#	
1	8F	61.83	6P	1.66	+++	6P	2.20	+++	7F	22.00	++	+++	0	0	6	
2	4F	69.33	3F	155.33	+ -	3P	10.00	+++	3P	9.00	+++	+++	1	2	3	
3	8F	24.14	9F	88.57	--	7P	8.71	+++	7P	10.41	+++	+++	8	2	7	
4	4P	14.75	4P	14.75	I	4P	0.00	+	4P	29.50	-	I	0	0	4	
5	1P	0.00	1P	0.00	I	1P	0.00	I	1P	0.00	I	I	0	0	1	
6	6F	67.60	5P	0.00	+++	5P	14.20	+++	5P	18.20	+++	+++	0	0	5	
7	3P	0.00	3P	0.00	I	3P	0.00	0	4F	19.33	---	I	0	0	3	
8	6P	6.33	6P	6.33	I	6P	3.33	+	6P	5.00	+	I	0	0	6	
9	9F	64.28	9F	64.28	I	7P	2.54	+++	7P	3.86	+++	+++	9	6	7	
10	5P	6.00	5P	21.20	-	5P	9.00	-	5P	28.20	-	-	0	0	5	
11	4F	234.00	3F	119.00	++	2P	17.00	+++	2P	17.00	+++	+++	1	1	2	
12	4P	0.75	4P	0.75	I	4P	7.25	-	4P	1.25	-	I	0	0	4	
13	4P	0.00	4P	0.00	I	4P	0.00	0	4P	0.00	0	I	0	0	4	
14	6P	0.00	6F	308.17	--	6P	0.00	0	7F	38.17	---	I	4	4	6	
15	6F	37.60	6F	33.40	+	5P	33.80	+++	6F	37.60	0	+	8	1	5	
16	2P	14.00	2P	14.00	I	2P	0.00	+	2P	14.00	0	I	0	0	2	
17	3F	114.50	2P	13.50	+++	2P	13.50	+++	2P	13.50	+++	+++	0	0	2	
18	3P	0.00	3P	0.00	I	3P	0.00	0	4F	1.67	---	I	0	0	3	
19	5P	0.00	5P	0.00	I	5P	0.00	0	6F	13.20	---	I	0	0	5	
20	5F	50.75	4F	1046.20	+ -	4P	0.00	+++	4P	6.25	+++	+++	1	3	4	
21	1P	0.00	1P	0.00	I	1P	0.00	I	1P	0.00	I	I	0	0	1	
22	1P	0.00	2F	79.00	---	1P	0.00	I	1P	0.00	I	I	1	1	1	
23	8F	62.67	8F	74.00	-	7F	22.33	++	7F	34.00	++	++	3	2	6	E*
24	9F	80.29	8F	574.00	+ -	7P	21.00	+++	7P	19.43	+++	+++	2	4	7	
25	4P	0.25	5F	633.00	---	4P	0.75	-	4P	20.25	-	-	6	6	4	
26	4P	12.00	4P	12.00	I	4P	0.00	+	4P	8.75	+	I	0	0	4	
27	5P	0.00	5P	0.00	I	5P	8.80	-	5P	11.60	-	I	0	0	5	
28	6F	119.25	6F	111.00	+	4P	0.00	+++	4P	4.75	+++	+++	3	4	4	
29	4P	0.00	4P	0.00	I	4P	0.00	0	4P	4.75	-	I	0	0	4	
30	6P	14.00	6P	14.00	I	6P	9.83	+	6P	9.83	+	I	0	0	6	
31	3P	1.33	3P	1.33	I	3P	0.00	+	3P	12.00	-	I	0	0	3	
32	5P	7.60	5P	7.60	I	5P	0.00	+	5P	0.00	+	I	0	0	5	
33	5F	72.00	4P	13.25	+++	4P	21.50	+++	4P	21.50	+++	+++	0	0	4	
34	4P	8.00	3F	70.25	+ -	4P	25.00	-	4P	15.50	-	-	2	2	4	
35	5P	22.80	5P	22.80	I	5P	24.00	-	5P	24.00	-	I	0	0	5	
36	5P	0.00	4F	264.80	+ -	5P	15.80	-	5P	10.20	-	-	4	2	5	
37	4P	0.00	4P	23.50	-	4P	4.50	-	4P	0.00	0	-	0	0	4	
38	3P	13.66	3P	13.66	I	3P	18.33	-	3P	18.33	-	I	0	0	3	
39	4P	0.00	5F	95.75	---	4P	0.00	0	4P	17.20	-	0	1	2	4	
40	5F	60.80	4P	35.50	+++	4P	0.00	+++	4P	29.25	+++	+++	0	0	4	
41	6F	45.20	5P	19.20	+++	6F	57.80	-	6F	179.00	-	+++	0	0	5	
42	6F	39.00	6F	33.00	+	5P	5.20	+++	5P	5.20	+++	+++	2	2	5	
43	5P	8.80	5P	0.20	+	5P	0.00	+	5P	0.00	+	+	0	0	5	
44	4P	10.50	4P	10.50	I	4P	19.50	-	4P	19.50	-	I	0	0	4	
45	8F	49.00	8F	520.25	-	9F	12.12	-	8P	33.75	++	++	10	4	8	
46	8F	91.17	5F	818.83	+ -	6P	18.00	+++	6P	7.50	+++	+++	4	4	6	
47	2P	0.00	2P	0.00	I	2P	0.00	I	3P	22.00	--	I	0	0	3	
48	3P	0.00	3P	0.00	I	3P	0.00	+	4F	13.33	---	I	0	0	3	
49	7F	95.67	5F	373.17	+ -	7F	55.83	+	7F	244.01	-	+ -	11	10	6	A*
50	5P	3.40	5P	3.40	I	5P	31.20	-	5P	9.20	-	I	0	0	5	
51	3F	45.00	3F	45.00	I	2P	0.00	+++	2P	0.00	+++	+++	3	4	2	
52	5F	41.25	4P	1.75	+++	4P	5.75	+++	5F	37.80	+	+++	0	0	4	
53	5F	33.33	4P	19.75	+++	4P	11.25	+++	4P	20.25	+++	+++	0	0	4	
54	4P	15.50	4P	15.50	I	4P	21.75	-	4P	17.75	-	I	0	0	4	
55	8F	260.00	5F	57.50	+++	5F	34.50	+++	6F	74.50	+++	+++	10	2	4	E
56	1P	0.00	1P	0.00	I	1P	0.00	I	1P	0.00	I	I	0	0	1	
57	5P	10.00	5P	10.00	0	5P	13.20	-	5P	13.20	-	0	0	0	5	
58	2P	0.00	3F	34.00	--	2P	0.00	I	2P	0.00	0	I	7	7	2	
59	6F	52.40	5F	120.80	+ -	5P	0.00	+++	5P	6.80	+++	+++	1	2	5	
60	3P	0.00	6F	168.00	---	4P	0.00	-	4P	22.25	--	-	1	2	4	
61	4P	0.50	6F	661.75	---	4P	0.00	+	4P	36.75	-	+	1	2	4	
62	5F	32.25	5F	82.00	--	5F	30.50	+	5F	31.25	+	--	10	10	4	E
63	4F	73.00	6F	223.33	--	3P	5.33	+++	3P	2.66	+++	+++	4	4	3	
	(a)		(b)			(c)			(d)				(e)	(f)	(g)	

* 'E' at the far right side of the table means that an 'exemption' was granted during the 1986 Census.
'A' means that an adjust (e.g., block split is needed to generate an acceptable solution.

(a) total districts = 296
'fails' = 26 F

(b) total districts = 284
% of 1986 districts = 96%
'fails' = 27 F

(c) total districts = 273
% of 1986 districts = 92%
'fails' = 6 F

(d) total districts = 277
% of 1986 districts = 94%
'fails' = 13 F

(e) number of steps for the dwelling based approach = 118
average number of steps = 1.9

(f) number of steps for the PFOM based approach = 95
average number of steps = 1.5

(g) target number of districts = 263

TABLE G.2 QUALITATIVE PERFORMANCE ASSESSMENT FOR DWELLING COUNT-BASED DISTRICTINGS (COMPARISONS WITH ACTUAL 1986 EAs)

1981 EAs — MANUAL

MANUAL	Pass	Fail
Pass	28	9
Fail	8	18

HH CADP — AUTOMATED

MANUAL	Pass	Fail
Pass	37	0
Fail	20	6

PFOM CADP — AUTOMATED

MANUAL	Pass	Fail
Pass	32	5
Fail	18	8

TOTAL — AUTOMATED

MANUAL	Pass	Fail
Pass	37	0
Fail	22	4

TABLE G.3 QUANTITATIVE PERFORMANCE ASSESSMENT FOR DWELLING COUNT-BASED DISTRICTINGS (COMPARISON WITH ACTUAL 1986 EAs)

Autodistricting verses Actual

METHOD	+++	++	+	0	I	+ -	-	--	---	Total	WTed Net	UnWTed Net
1981 EAs	9	1	4	1	26	8	4	5	5	63	4	0
HH CADP	21	1	12	7	6	0	16	0	0	63	77	18
PFOM CADP	18	3	7	5	4	0	19	2	5	63	31	2
SCORE1	21	2	3	2	27	1	6	1	0	63	62	19

These summary tables (G.2 and G.3) have been compiled from Table G.1, and are replicated in Chapter 5.

TABLE G.4 METHOD COMPARISONS BASED ON PFOM SCORES

CTP#	1986 EAs	1981 EAs	+/-	HH CADP	+/-	PFOM CADP	+/-	SCORE	STEPS		#	
1	7P 0.0	6P 83.4	+ -	6P 146.3	+ -	7P 57.86	-	+ -	0	0	7	
2	4F 521.0	3F 2375.0	+ -	3P 0.0	+++	3P 0.00	+++	+++	1	2	3	
3	8F 28.0	9F 91.7	--	7P 112.4	++	7P 0.00	+++	+++	8	2	7	
4	4P 10.0	4P 9.6	+	4P 0.6	+	4P 45.00	-	+	0	0	4	
5	1P 0.0	1P 0.0	I	1P 0.0	I	1P 0.00	I	I	0	0	1	
6	6F 144.0	5P 0.0	+++	5P 0.0	+++	5P 0.00	+++	+++	0	0	5	
7	4P 0.0	3P 0.0	+	3P 0.0	+	4P 0.00	0	+	0	0	4	
8	6P 0.0	6P 0.0	0	6P 0.0	0	6P 0.00	0	0	0	0	6	
9	9F 175.0	9F 225.0	-	7P 2.6	+++	7P 0.00	+++	+++	9	6	7	
10	5P 0.0	5P 30.6	-	5P 15.4	-	5P 71.60	-	-	0	0	5	
11	4F 404.0	3F 445.0	+ -	2P 9.5	+++	2P 9.50	+++	+++	1	1	2	
12	4P 0.0	4P 0.0	0	4P 0.0	0	4P 0.00	0	0	0	0	5	
13	4P 0.0	4P 0.0	0	4P 0.0	0	4P 0.00	0	0	0	0	4	
14	6F 307.0	6F 401.1	-	6F 619.8	-	7P 45.14	+	+	3	4	7	
15	6P 73.0	6P 72.5	+	5F 50.0	--	6P 50.50	+	+	0	0	6	
16	2P 0.0	2P 0.0	0	2P 0.0	0	2P 0.00	0	0	0	0	2	
17	3F 472.0	2P 37.0	+++	2P 37.0	+++	2P 37.00	+++	+++	0	0	2	
18	3P 5.0	3P 3.5	+	3P 0.0	+	4P 0.00	-	-	0	0	4	
19	5P 9.0	5P 8.1	+	5P 50.0	-	6P 0.00	-	-	0	0	6	
20	5F 112.0	4F 1913.0	+ -	4P 55.6	+++	4P 0.00	+++	+++	1	3	4	
21	1P 0.0	1P 0.0	I	1P 0.0	I	1P 0.00	I	I	0	0	1	
22	1P 0.0	2F 46.0	---	1P 0.0	I	1P 0.00	I	I	1	1	1	
23	8F 300.0	8F 157.0	+	7F 322.4	+ -	7P 44.28	+++	+++	10	2	7	A
24	9F 176.0	8F 259.0	+ -	7F 228.1	+ -	7P 6.00	+++	+++	2	4	7	
25	4F 2904.0	5F 2044.0	+ -	4F 1612.0	+	4P 0.00	++	++	6	6	4	
26	4P 102.0	4P 104.0	-	4P 0.0	+	4P 0.00	+	-	0	0	4	
27	5P 43.0	5P 45.0	-	5P 0.0	+	5P 0.00	+	-	0	0	5	
28	6F 287.0	6F 393.0	-	4P 77.8	+++	4P 0.00	+++	+++	3	4	4	
29	4P 95.0	4P 95.0	0	4P 0.1	+	4P 0.00	+	0	0	0	4	
30	6P 40.0	6P 40.0	0	6P 20.0	+	6P 20.00	+	0	0	0	6	
31	3P 0.0	3P 0.0	0	3P 0.0	0	3P 0.00	0	0	0	0	3	
32	5P 7.8	5P 7.8	0	5P 0.0	+	5P 0.00	+	I	0	0	5	
33	5F 141.0	4P 32.0	+++	4P 36.0	+++	4P 0.00	+++	+++	0	0	4	
34	4P 0.0	3P 17.1	+ -	4P 82.8	-	4P 43.50	-	-	0	0	4	
35	5P 111.2	5P 111.2	I	5P 72.0	+	5P 72.00	+	I	0	0	5	
36	5P 0.0	4F 291.0	+ -	5P 26.2	-	5P 25.80	-	+ -	4	2	5	
37	4P 0.0	4P 61.0	-	4P 0.0	0	4P 0.00	0	0	0	0	4	
38	3P 15.0	3P 15.0	0	3P 39.3	-	3P 39.33	-	0	0	0	3	
39	4P 30.0	5F 648.0	---	4P 44.6	-	4P 0.00	+	+	1	2	4	
40	5F 990.0	4F 19040.0	+ -	4F 1163.0	+ -	5F 134.25	+	+	2	10	4	E
41	6F 44.0	5F 266.0	+ -	6F 179.0	-	6F 107.80	-	+ -	10	10	5	E
42	6F 672.6	6P 107.0	+++	5P 35.6	+++	5P 0.00	+++	+++	0	0	5	
43	5P 40.0	5P 54.0	-	5P 1.0	+	5P 0.00	+	-	0	0	5	
44	4P 29.0	4P 30.0	-	4P 21.0	+	4P 21.00	+	-	0	0	4	
45	8P 11.0	8F 487.0	--	8P 31.6	-	8P 0.00	+	+	4	4	8	E
46	8F 2067.0	5F 88823.0	+ -	6P 127.5	+++	6P 1.83	+++	+++	4	4	6	
47	2F 175.0	2F 175.0	0	2F 262.5	-	3P 42.33	+	+	2	2	3	
48	3P 0.0	3P 0.0	0	3P 0.0	0	4P 0.00	--	0	0	0	4	
49	7F 151.0	5F 905.0	+ -	7F 244.0	-	7F 123.00	+	+ -	10	10	6	A
50	5P 0.0	5P 0.0	0	5P 63.3	-	5P 23.00	-	0	0	0	5	
51	3F 64.5	3F 64.5	0	2F 183.3	+ -	2P 0.00	+++	+++	3	4	2	
52	5P 6.0	4P 0.0	++	4F 390.8	---	5P 0.00	+	++	0	0	5	
53	5F 7.0	4P 0.0	+++	4F 1007.0	+ -	4P 0.00	+++	+++	0	0	4	
54	4F 415.0	4F 415.0	I	4P 95.3	++	4P 0.00	++	++	2	4	4	
55	8F 928.6	5F 502.0	+ -	5P 29.4	++	5F 238.80	+ -	++	10	2	5	E
56	1P 0.0	1P 0.0	I	1P 0.0	I	1P 0.00	I	I	0	0	1	
57	5P 11.4	5P 11.4	I	5P 29.4	-	5P 29.40	-	I	0	0	5	
58	2P 0.0	3P 95.5	--	2P 0.0	0	3F 167.30	---	0	0	0	3	
59	6F 115.0	5F 246.6	+ -	5P 0.0	+++	5P 0.00	+++	+++	1	2	5	
60	3P 14.0	6F 294.0	---	4P 0.0	+ -	4P 0.00	+ -	+ -	1	2	4	
61	4P 118.0	6F 282.0	---	4P 0.6	+	4P 0.00	+	+	1	2	4	
62	5P 42.0	5F 588.0	--	5P 25.0	+	5F 274.60	--	+	4	10	5	E
63	4F 113.0	6F 518.0	---	3F 554.1	+ -	3P 0.00	+++	+++	4	4	3	
	(a)	(b)		(c)		(d)			(e)	(f)	(g)	

* 'E' at the far right side of the table means that an 'exemption' was granted during the 1986 Census.
'A' means that an adjust (e.g., block split is needed to generate an acceptable solution.

(a) total districts = 296
'fails' = 25 F

(b) total districts = 284
% of 1986 districts = 96%
'fails' = 27 F

(c) total districts = 267
% of 1986 districts = 90%
'fails' = 13 F

(d) total districts = 278
% of 1986 districts = 94%
'fails' = 6 F

(e) number of steps for the
dwelling based approach = 118
average number of steps = 1.8

(f) number of steps for the
PFOM based approach = 109
average number of steps = 1.7

(g) target number of districst = 276

TABLE G.5 QUALITATIVE PERFORMANCE ASSESSMENT FOR FIGURE OF MERIT-BASED DISTRICTINGS (COMPARISONS WITH ACTUAL 1986 EAs)

1981 EAs — MANUAL

MANUAL	Pass	Fail
Pass	31	7
Fail	5	20

HH CADP — AUTOMATED

MANUAL	Pass	Fail
Pass	36	2
Fail	14	11

PFOM CADP — AUTOMATED

MANUAL	Pass	Fail
Pass	36	2
Fail	21	4

TOTAL — AUTOMATED

MANUAL	Pass	Fail
Pass	38	0
Fail	22	3

TABLE G.5 QUANTITATIVE PERFORMANCE ASSESSMENT FOR FIGURE OF MERIT-BASED DISTRICTINGS (COMPARISON WITH ACTUAL 1986 EAs)

Autodistricting verses Actual

METHOD	+++	++	+	0	I	+ -	-	--	---	Total	WTed Net	UnWTed Net
1981 EAs	5	1	6	13	6	14	9	4	5	63	- 9	- 6
HH CADP	11	3	14	8	4	8	13	1	1	63	35	13
PFOM CADP	17	2	17	7	4	2	11	2	1	63	53	19
SCORE1	17	4	10	12	17	5	8	0	1	63	58	14

The analysis of these results is reported in Chapter 5.